PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07K 13/00, C12N 9/48, 9/64 — A1

(11) International Publication Number: WO 91/10685

(43) International Publication Date: 25 July 1991 (25.07.91)

(21) International Application Number: PCT/US91/00340

(22) International Filing Date: 17 January 1991 (17.01.91)

(30) Priority data:
467,880 19 January 1990 (19.01.90) US

(71) Applicant: SERAGEN, INC. [-/US]; 97 South Street, Hopkinton, MA 01748 (US).

(72) Inventors: BLEACKLEY, Robert, C. ; 9114 117th Street, Edmonton, Alberta T6G 1R9 (CA). LOBE, Corrine, G. ; 13008 62nd Street, Edmonton, Alberta P5A 0V2 (CA). PAETKAU, Verner, H. ; 20110805 79th Avenue, Edmonton, Alberta P6E 1S6 (CA). JAMES, Michael, N., G. ; 8347 120th Street, Edmonton, Alberta P6G 1X1 (CA). MURPHY, Michael ; 10979 35A Avenue, Edmonton, Alberta T6J 0A2 (CA).

(74) Agent: CLARK, Paul, T.; Fish & Richardson, One Financial Center, Suite 2500, Boston, MA 02111-2658 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: CYTOTOXIC CELL-SPECIFIC PROTEASE-RELATED MOLECULES AND METHODS

(57) Abstract

A peptide capable of inhibiting, in a mammal to which the molecule is administered, the biological activity of a cytotoxic cell protease.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

CYTOTOXIC CELL-SPECIFIC PROTEASE-RELATED MOLECULES AND METHODS

Background of the Invention

This is a continuation-in-part of copending application serial number 002,960 filed on January 13, 1987.

This invention relates to protease inhibitors.

Thymus derived (T) lymphocytes play a major role in the immune system. Maturation of the T cell lineage involves three distinct stages: (a) generation of a T cell precursor from a pluripotent stem cell, (b) differentiation in the thymus, and (c) migration of mature cells to the peripheral tissues. Maturation of T cells within the thymus is antigen independent. However, once they have left the thymus, upon interaction with an antigen they are driven through the final steps of differentiation to become mature cells. These final steps are complex and involve interactions with other cells and soluble effector molecules.

Several subsets of T cells have been recognized among activated peripheral T cells. There are three main classes: helper, suppressor, and cytotoxic. Helper T lymphocytes potentiate immune responses (both humoral and cell-mediated) either by cell-cell contact or by synthesis and secretion of factors. These factors, although synthesized in response to an antigen-specific signal, can be either antigen-specific or antigen-nonspecific. Suppressor T lymphocytes, inhibit the functions of other lymphocytes, again either directly or via soluble factors. Cytotoxic T lymphocytes are the effector cells in cell mediated immune reactions. They specifically recognize foreign antigens on the surface of cells, bind to them, and cause the target cell to lyse. Cytotoxic T lymphocytes are

known to cause or exacerbate autoimmune diseases such as rheumatoid arthritis, and are also involved in allograft rejection and graft-versus-host disease.

The various steps in the process of cytotoxic T lymphocyte induced lysis have been analyzed in some detail, e.g., Berke, (1983) Immunol. Rev. 72:5; Nabholz & MacDonald, (1983) Ann. Rev. Immunol. 1:273. Recent studies by Padack & Konigsberg, (1984) J. Exp. Med. 160:695 and Henkart et al., (1984) J. Exp. Med. 160:75 have suggested that the dense cytoplasmic granules seen in CTL and natural killer cells are directly involved in target cell lysis by a mechanism involving transmembrane channels.

A general description of cytotoxic T lymphocytes, natural killer cells, and killer (K) cells is contained in Stites et al., Basic & Clinical Immunology 227-31 (Lange Medical Publications, Los Altos, Ca., 1984).

Summary of the Invention

In general, the invention features a vector containing a DNA sequence encoding the CCP1 protein.

In another aspect the invention features a vector containing a DNA sequence encoding the CCP2 protein.

In another aspect the invention features a vector containing a DNA sequence encoding the hCCP1 protein.

In another aspect the invention features a vector containing a DNA sequence encoding the hCCPX protein.

In another aspect the invention features substantially pure CCP1 protein expressed from a vector containing a DNA sequence encoding the CCP1 protein. Substantially pure means a preparation with a purity of 95% or greater by weight, and free of the proteins, lipids, and carbohydrates with which the protein is naturally associated.

In another aspect the invention features substantially pure CCP2 protein expressed from a vector containing a DNA sequence encoding the CCP2 protein.

In another aspect the invention features substantially pure hCCP1 protein expressed from a vector containing a DNA sequence encoding the hCCP1 protein.

In another aspect the invention features substantially pure hCCPX protein expressed from a vector containing a DNA sequence encoding the hCCPX protein.

In another aspect, the invention features a peptide of the formula: Asp-Val-Asp-Ala; Ala-Pro-Asp-Ala; Ala-Asn-Pro-Ala; Phe-Pro-Arg-Phe; Ala-Pro-Arg-Phe; Phe-Pro-Asp-Phe; Phe-Pro-Asn-Phe; Phe-Asn-Pro-Phe; or Phe-Asp-Pro-Phe.

The term competitive inhibition, as used herein, refers to inhibition in which the inhibitor combines with the free protease such that it competes with the normal substrate of the protease. Competitive inhibition is described, e.g., in Lehninger, Biochemistry 197-200 (Worth, 2d ed. 1975).

The term protease, as used herein, refers to an enzyme that hydrolyzes, and thus cleaves, peptide bonds.

Cytotoxic lymphocytes, e.g. cytotoxic T lymphocytes (sometimes called T killer cells) and natural killer cells are described in Jandl, Blood: Textbook of Hematology (Little, Brown and Co., Boston, 1987) hereby incorporated by reference.

The term serine protease, as used herein, refers to a protease which has a serine residue at the active site of the enzyme.

The term peptide, as used herein, includes proteins as well as peptides too short to be characterized as proteins. Generally those peptides having a molecular weight of greater than 5,000 are characterized as proteins.

The term cytotoxic cell protease, as used herein, refers to any protease, preferably a serine protease, that has 30% or greater homology, more preferably 50% or greater homology, with the protein encoding sequence of the murine C11 gene, and which cleaves at different sites than does plasmin. Preferably the cytotoxic cell protease is expressed by cytotoxic lymphocytes, more preferably exclusively by cytotoxic lymphocytes.

Cytotoxic lymphocytes produce, as part of their cytotoxic activity, proteases, some of which, we have discovered, cleave proteins at sites different from the sites cleaved by proteases such as plasmin produced by other cells of the body. These proteases are members of the cytotoxic cell protease family. The inhibitory molecules of the invention, since they mimic the unique cleavage sites recognized by cytotoxic cell proteases, can exclusively inhibit cytotoxic cell proteases e.g., those produced by cytotoxic lymphocytes. Thus a person suffering from an immune disorder, or experiencing allograft rejection, can be administered a molecule of the invention to inhibit the cytotoxic lymphocytes involved in the disease or rejection process, and the administered molecule will not interfere with, for example, lysis of blood clots, or other normal protease-dependent functions.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments and from the claims.

Description of the Preferred Embodiment

The structure, synthesis, and use of the preferred embodiments are discussed next, after the drawings are briefly described.

Drawings

Fig. 1 is a graph showing the correlation of a protease mRNA expression (·) with cell activation in a mixed lymphocyte culture.

Fig. 2 is a partial nucleotide sequence comparison of two protease-encoding cDNA's.

Fig. 3 is the nucleotide sequence of one of said cDNA's and the predicted protein structure it encodes.

Fig. 4 is a partial amino acid sequence comparison of five serine proteases.

Fig. 5 is the sequence of CCP2.

Fig. 6 is the sequence of hC11, the human analog of the murine C11 gene.

Fig. 7 is a restriction map of the hCCPX gene.

Fig. 8 is the nucleotide sequence of the hCCPX gene.

Fig. 9 is the predicted cDNA sequence encoded by the hCCPX gene.

Fig. 10 is the amino acid sequence of proteins encoded by hCCPX and the CCP genes.

Fig. 11 is the amino acid sequence of some protease inhibitors of the invention.

Table 1 shows the expression of C11 mRNA in infiltrating cells of tissue grafts.

Table 2 shows the degree of homology between CCP1 and various proteins.

Table 3 shows the effect of peptides of the invention on the cytotoxicity of cells from a cyclosporine-A mixed lymphocyte reaction.

Table 4 shows the effect of peptides of the invention on the cytotoxicity of cytotoxic T-cells activated with ConA and interleukin 2.

The Appendix is a copy of Murphy et al. (1988) Proteins: Structure, Function, and Genetics 4:190-204 which

provides a detailed example of computer aided analysis of enzyme and substrate structure.

Structure

The inhibitory molecules of the invention competitively inhibit the activity of cytotoxic cell proteases produced e.g. by cytotoxic lymphocytes, while not inhibiting the activity of proteases produced by other cell types or any other proteases produced by the cells producing the cytotoxic cell proteases. Preferably the inhibitory molecules are peptides.

Cytotoxic lymphocytes synthesize a characteristic set of cytotoxicity-related proteases which are expressed at much reduced levels, if at all, in other subsets of lymphocytes. The cytotoxicity- related proteases can be divided into two groups, effector proteases and non-effector proteases. Effector proteases are released by a cytotoxic lymphocyte when it comes in contact with a target cell, and break down proteins in the membrane of the target cell or enter the target cell and hydrolyze intracellular proteins, leading to the cell's destruction. Non-effector proteases are involved in the enzymatic processes that lead to the production and/or release of the effector proteases (or other effector molecules) from the lymphocyte. Inhibiting the action of either an effector protease or a non-effector protease inhibits the ability of cytotoxic lymphocytes to destroy a target cell.

The preferred peptides contain the two amino acids that constitute the cleavage site recognized by the protease, and have between 3 and 20 (more preferably between 3 and 5) amino acids residues. Shorter peptides are preferred because they are, in general, readily taken up cells. The peptides should not contain a cleavage site recognized by other proteases, for example, those sites

described by Zreighton, Proteins: Structure and Molecular Properties 336-37, 427-38 (W.H. Friedman, N.Y., 1983).

Described in Example 1 below is the isolation, cloning, and characterization of two genes expressed exclusively in the cytotoxic T lymphocytes of mice. (Exclusively means that either the genes are not expressed, or are only expressed in very low (less than 5 molecules of mRNA per cell) levels, in other types of cells in the organism). Example 2 describes the sequencing of the two genes, the determination of the amino acid sequence of the protease which one of the genes encodes, and the characterization of the protease. Example 3 describes the identification and isolation of a human gene (hCII) encoding a cytotoxic cell protease (hCCP1) produced exclusively by human cytotoxic T lymphocytes. Example 4 describes the isolation, cloning, and characterization of a gene encoding another human cytotoxic cell protease, human cytotoxic cell protease X (hCCPX). Example 5 describes the sequencing of the hCCPX gene, the determination of the amino acid sequence of the hCCPX protease, and the characterization of the protease. Example 6 describes the determination of three dimensional structure of a cytotoxic cell protease and the structure of a peptide that can act as a competitive inhibitor of that protease. Example 7 describes several inhibitors of the invention. Example 8 describes the production of substantially pure proteases and their use in the design of inhibitors.

Example 1

Cells - The cytotoxic T-cell lines MTL2.8.2 and MTL11.1 were generated from CBA/J mice as described by Bleackley et al., (1982) J. Immunol. 128:758. EL4.E1 is an interleukin 2 (IL-2)-producing variant of the EL4 cell line described by Farr et al., (1980) J. Immunol. 125:2555. CH1

is a CBA/J α CBA/J X BALB/c antigen- specific helper T-cell line. It was produced from a 2-day mixed lymphocyte culture by continuous restimulation with irradiated F_1 spleen cells in RPMI 1640 medium supplemented with 10% fetal bovine serum and 100 μM 2-mercaptoethanol (RHFM). To generate human cytotoxic T lymphocytes (CTL), peripheral blood lymphocytes were incubated in RHFM and stimulated with irradiated allogeneic cells at days 0 and 7 and harvested at day 10. The fetal-derived cells used are described by Teh et al., (1985) J. Immunol. 135:1582. For the time course of cell activation, spleen cells from CBA/J mice were incubated in RHFM (10^6 cells per ml) and purified IL-2 (described by Riendeau et al., (1983) J. Biol. Chem. 258:12114), either with an equal number of mitomycin C-treated EL4.E1 cells or Con A (2 μg/ml). Samples were removed at day 1 through day 6, assayed for cytotoxic activity by the procedure described in Shaw et al., (1978) J. Immunol. 120:1974, and analyzed by cytodot hybridization.

cDNA Library Construction - Double-stranded cDNA was synthesized from 4 μg of MTL.2.8.2 mRNA as described by Gubler and Hoffman, (1983) Gene 25:263. Following repair with the Klenow fragment of DNA polymerase and T4 DNA polymerase to maximize flush ends, phosphorylated EcoRI linkers (P-L Biochemicals) were ligated to the cDNA in 70 mM Tris-HCl, pH 7.6/10 mM $MgCl_2$/5 mM dithiothreitol/1 mM ATP/1 unit of T4 DNA ligase at 14°C overnight (Goodman & MacDonald, (1979) Methods Enzymol. 68:75). After digestion with EcoRI, the product was run on a 5-ml Sepharose 4B column, and the excluded fractions were pooled and ethanol-precipitated. The cDNA was ligated to EcoRI/bacterial alkaline phosphatase-treated pUC13 (P-L Biochemicals) in 66 mM Tris-HCl, pH 7.6/6.6 mM $MgCl_2$/10 mM dithiothreitol/1 mM ATP. Reactions were heated to 37°C for 5 min, quick-chilled

before the addition of 1 unit of T4 DNA ligase, and incubated at 14°C for 2 hr. Escherichia coli JM83 cells were made competent by using the $CaCl_2$/RbCl procedure described by Maniatis et al. in Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982) and were transformed with the ligated cDNA. White colonies (those containing inserts) were ordered in 96-well microtiter plates and stored in LB medium containing 20% glycerol at -70°C.

Differential Screening - Colonies were replicated in triplicate onto nitrocellulose filters, grown for 6 hr, and then amplified on chloramphenicol (100 μg/ml) for 12 hr. Bacteria were lysed, and the filters were prewashed to remove bacterial debris, as described by Maniatis, supra. Prehybridization at 42°C for 12-20 hr was done in 50% (vol/vol) formamide containing 2x Denhardt's solution (1x Denhardt's solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin), 4x SET buffer (1x SET buffer = 0.6 M NaCl/0.12 M Tris-HCl, pH 8/1 mM EDTA), 0.1% $NaDodSO_4$, 100 μg of yeast tRNA per ml, and 125 μg of poly(A) per ml (Sigma). Hybridization in the same buffer included 1-5 x 10^5 cpm of cDNA probe per ml synthesized from mRNA with 20 μg of T-primers per ml (Collaborative Research, Waltham, MA); 50 mM Tris-HCl (pH 8.3); 10 mM $MgCl_2$; 5 mM dithiothreitol, 500 μM each of dGTP, dATP, and dTTP; 70 mM KCl; 30 μCi (1 Ci = 37 GBq) of [α-^{32}P]dCTP (New England Nuclear, 300 Ci/mmol); and 15 units of avian myeloblastosis virus reverse transcriptase at 42°C for 60 min. Template RNA was hydrolyzed by the addition of NaOH to 1.5 M. Samples were boiled for 3 min and fractionated by Sephadex G-50 column chromatography. Filters were washed in 5x SET buffer for 15 min at 22°C and then in 2x SET buffer/50% formamide for 20 min at 42°C and were exposed to film (Kodak

X-Omat AR) with an intensifying screen for 1 to 3 days at -70°C. Hybridized probe was removed by boiling the filters for 10 min in distilled water.

Blot Analysis - Cytodots were prepared as described by White and Bancroft (1982) J. Biol. Chem. 257:8569. For blot-hybridization analysis, total cytoplasmic RNA (10 μg) or poly(A)$^+$ mRNA (2 μg) was denatured in 6.3% formaldehyde/50% formamide at 55°C and size fractionated on a 0.8% agarose gel containing 0.66% formaldehyde. RNA was transferred to nitrocellulose as described by Thomas (1980) Proc. Natl. Acad. Sci. USA 77:5201. Plasmid DNA was digested with EcoRI, run on a 0.7% agarose gel, and transferred to nitrocellulose, as described by Southern (1975) J. Mol. Biol. 26:365. Filters were baked at 80°C for 2 hr, then prehybridized at 42°C for 6-12 hr in 50% formamide containing 20 mM phosphate buffer (pH 6.8), 2 mM pyrophosphate, 100 μM ATP, 5x Denhardt's solution, 0.75 M NaCl, 0.075 M sodium citrate (pH 7), 100 μg of salmon sperm DNA per ml, 0.1% $NaDodSO_4$, 50 μg of poly(A) per ml, and 2.5 mM EDTA. Hybridization was carried out in the same buffer with a nick-translated plasmid of specific activity 1×10^8 cpm/μg (Bethesda Research Laboratories kit) at 1×10^6 cpm/ml.

Results - Triplicate copies of the library were hybridized first with cDNA synthesized from MTL2.8.2 mRNA, then, after autoradiography and washing, with helper T-cell cDNA, and finally with thymocyte cDNA. Colonies that gave a higher hybridization signal with killer cell mRNA in at least two of the three copies of the library were picked. Upon rescreening, again in triplicate, 36 of these 121 colonies appeared to be clearly CTL-specific. Plasmid DNA isolated from these colonies was cut with EcoRI, and a series of cross-hybridizations was performed. Two clones

were chosen for more extensive analysis: clone B10 because it appeared to be the most abundant in the library, cross-hybridizing strongly with eight other inserts, and clone C11 because it weakly cross-hybridized with B10 but not with all B10-related clones (one other C11-related sequence was found).

Cytodots prepared from a variety of cells and tissues were hybridized with nick-translated B10 and C11. The number of cells per dot was 10^4. The data with probe C11 are similar and are not discussed. The highest signal was detected in MTL2.8.2--i.e., the killer cell line that was used to generate the cDNA library. A weaker but positive signal was observed with MTL-III, a variant of MTL2.8.2 that had a low level of cytotoxicity and had become IL-2 and antigen independent. A similar level of expression was observed in a novel T-cell clone derived from murine fetal thymus of Teh, supra. In all over 20 cytotoxic T cell lines and cultures have been tested and all have been positive for B10 and C11 expression.

Natural killer (NK) and T killer (TK) cells were purified, cultured, and tested for the expression of C11 mRNA by the methods described in Manyak et al. (1989) J. Immunol. 142:3707-3713. Culturing NK cells in IL-2 induced: i) lytic activity, ii) chymase and tryptase enzymatic activities and iii) the total mRNA levels of the C11 gene in a dose-dependent manner. C11 mRNA reached peak activity on days 5 to 7 of culture. Similar results were seen with TK cells.

There was no evidence for expression of B10 or C11 in either mouse thymocytes or a helper T-cell line (CH1) that secretes IL-2 in response to antigen. Mouse brain, mouse liver, and a human CTL line were similarly negative under the high-stringency conditions of this experiment. In

addition, no evidence for expression of B10 or C11 was found in a helper T-cell hybridoma that secretes an antigen-specific factor (Kwong et al. (1984)) J. Immunol. 133:653. To ensure that the negative samples did contain hybridizable RNA, all of the cytodots were reprobed with either a lymphocyte-specific probe or oligo(dT) or the T-cell antigen receptor ß-chain gene (Hendrick et al. (1984) Nature 308:153). Although the level of signal varied, all samples were positive.

To enrich for the B cells of a spleen cell suspension, lymphocytes were separated from adherent cells on Petri dishes and then treated with anti-Thy-1.2 antiserum. The enriched B cells were then incubated with lipopolysaccharide (LPS) or Con A or RHFM medium. After 24 hr, the cells were harvested, cytodots were prepared and the filter was probed with B10 or C11. No expression of either sequence could be detected in any sample. However, when the blot was hybridized with an immunoglobulin μ heavy chain probe (Calame et al. (1980) Nature 284:452) a strong positive signal was seen in the LPS-stimulated cells.

Poly(A)$^+$ RNA was isolated from a variety of cell sources, run on a denaturing agarose gel, and transferred to nitrocellulose. The same filter was probed first with nick-translated B10, then with C11, and finally with probe 10, a cloned gene that detects mRNA in a variety of cell types (Paetkau et al., in Contemporary Topics in Molecular Biology 10:35 (S. Gillis ed., Plenum, N.Y., 1984)). Probe B10 detected a single band (approximately 900 bases) in two different murine cytotoxic T cell clones, MTL2.8.2 and MTL11.1. No bands were detected in RNA from thymocytes, an antigen-specific helper cell line, or murine thymoma EL4. When the blot was reprobed with C11, again only the two cytotoxic T cell clones showed bands. However, in contrast

to B10, this probe hybridized to two bands, one of approximately 900 bases and the other of 1200 bases. Probe 10 detected a band in all cell samples. In addition, blot-hybridization analysis was performed on poly(A)$^+$ RNA from a number of murine cells including eleven CTL lines, two helper lymphocyte lines, brain cells, liver cells, three helper T-cell lines, unstimulated and LPS-stimulated B lymphocytes, and one B-cell myeloma. Of these, only the actively cytotoxic T cells expressed mRNAs that hybridized with B10 and C11. To ensure that all tracks contained hybridizable RNA, the blot was rehybridized with probe 10. A band of the expected size was seen in all tracks.

The results from the cytodots and blot-hybridization analysis indicates that both B10 and C11 are murine cytotoxic T lymphocyte specific.

CBA/J (H-2^k) spleen cells were stimulated with either mitomycin C-treated EL4 cells (Fig. 1A) or Con A (Fig. 1B). On each of the 6 days after stimulation, the level of cytotoxicity was measured in a chromium-release assay against EL4 (H-2^b) (□), S194(H-2^d) (Δ), and RI(H-2^k) (0) cell lines. Cytodots were also prepared on each of these days, and the blots were hybridized with nick-translated B10 and C11. Data are presented only for B10, as C11 gave indistinguishable results. Relative B10 mRNA levels (·) were determined by scanning densitometry on an ELISA plate reader. In the allo-specific response (Fig. 1A), the peak of cytotoxicity was observed on day 4, while the peak of B10 or C11 mRNA expression appeared to be on days 3 and 4. The peak of killing activity in the Con A-stimulated cells (Fig. 1B) was also at day 4; however, the peak of mRNA expression was very sharply on day 3. In both experiments, the mRNA expression was reduced to background levels by day 6, while there were still significant levels

of cytotoxicity on this day. When the cytodots were hybridized with ^{32}P-end- labelled oligo(dT), the peak of total mRNA was seen on day 2. (data not shown)

The experimental results illustrated in Fig. 1 indicates that the maximum expression of B10 and C11 mRNAs precedes the peak of cytotoxicity in an in vivo allogenic or mitogen-induced cytotoxic response by 24hrs; thus, they both fulfill the primary prerequisite for genes encoding proteins that are important in the lytic process.

In situ hybridization experiments indicate that a high proportion of T lymphocytes that infiltrate incompatible heart allografts in vivo express the C11 gene. Complete details of the in situ hybridization procedure, and all related techniques, are found in Mueller et al., (1988) J. Exp. Med. 167:1124-1136,

Transplantations in these experiments were performed as described in Mueller et al. (1988) J. Exp. Med. 167:1124-1136 and Billingham et al. (1977) Transplantation 23:171. In short, the myocardium of newborn (12-36 h) BALB/cJ (H-2^d) donor mice were diced into 0.1-0.2-cm fragments and subsequently transplanted under the kidney capsule of adult (6-8 wk) sex-matched C57 Bl/Ka recipients (H-2^b; experimental animals). As a control, adult BALB/cJ (H-2^d) mice received grafts from the same donor animals under the kidney capsule. On days 2, 4, 6, 8, 10, and 12 after transplantation, three experimental and two control animals were killed and 5-μm frozen sections through the graft were prepared. Labelled probe for in situ hybridization was prepared as described in Mueller et al. (1988) J. Exp. Med. 167:1124-1136 and as follows.

A 1.1-kb fragment of the C11 gene was subcloned into the polylinker of the transcription vector pSPT 672 using standard techniques. This vector ha a SP6 and a T7 promotor

at the 5' and 3' end of the multicloning site, respectively. After linearization of the vector with an appropriate restriction enzyme, sense and antisense probes were prepared using SP6-polymerase and T7-polymerase (both from New England Biolab, Beverly, MA) reactions and (S-35) UTP No. SJ 1303, Amersham Corp., Arlington Heights, IL) at a final concentration of 12 μM. The labelled nucleotide was dried down before adding the other reagents of the reaction mixture. A typical reaction (35 μl) contained 7 μl 5X SP6 buffer (final concentration; 40mM Tris-HCl, pH 7.9; 5mM $MgCl_2$; 2 mM spermidine); 3.5 μl 100 mM dithiothreitol (DTT); 3.5 μl ribonucleotides (CTP, ATP, and GTP; 10 mM each, in 10 mM Hepes, pH 7.4); 3.5 μl bovine serum albumin (BSA), 5 mg/ml; 1 μl Rnasin, 40 U/μl (New England Biolab); 1 μl linearized DNA template, 1 μg/μl; 13.5 μl H_2O. SP6 and T7 reactions were incubated for 90 min at 40°C and 37°C, respectively. DNA template was digested with DNase I (2U/μg DNA; Worthington) for 15 min at 37°C. The RNA probe was subsequently extracted with phenol/chloroform, separated on a Bio-Gel P-60 spin column, and ethanol precipitated after adding 7.5 μg of yeast tRNA per 10^6 cpm-labelled probe. The probe was subsequently resuspended at 2 X 10^9 cpm/μl in Tris-EDTA (TE), boiled for 2 min. and stored frozen at -70°C. For the hybridization, this probe was mixed with formamide (final concentration 50%), dextran sulfate (10%), DTT (100mM), NaCl (300mM), Tris-HCl, pH 7.5 (20mM), EDTA (5mM) Denhardt's solution (1X) at a concentration of 2 X 10^6 cpm/μl hybridization solution.

In situ hybridizations were performed according to Angerer et al. (1987) In In Situ hybridization: Applications to the CNS, K. Valentino, J. Eberwine, and J. Barchus, eds. Oxford University Press, New York pp. 42-70 as modified in Mueller et al. (1988) J. Exp. Med.

167:1124-1136. 5-μm-thick cryostat sections were placed on poly-L-lysine (Sigma Chemical Co.)-coated glass slides and fixed in 4% paraformaldehyde dissolved in 1X phosphate buffered saline (PBS) for 20 min., rinsed in PBS, and dehydrated through graded ethanol. Slides were stored at this stage at 4°C before being used for in situ hybridization. In situ hybridizations on different cell populations were done on sorted cells that were spun onto poly-L-lysine-coated glass slides with a Shandon cytocentrifuge. These cytospin preparations were fixed and hybridized as described for cryostat sections. The fixed sections or cytospin preparations were treated with proteinase K (Boehringer Mannheim, Federal Republic of Germany), 1 μg/ml in 100 mM Tris-HCl, pH 8.0, and 50 mM EDTA at 37°C for 30 min. The slides were postfixed again with 4% paraformaldehyde for 20 min. Free amino groups on tissue sections were acetylated by treatment with 0.25% acetic anhydride in 0.1 M triethanolamine for 10 min. For the hybridization step, 10 μl of the hybridization solution (described above) containing 10^6 cpm S-35 UTP-labelled RNA probe were placed on each section, covered with a siliconized coverslip (18 x 18 mm), and sealed with rubber cement. The sections were hybridized at 46°C for 16-18 h. Thereafter, the slides were washed in a solution containing 50% formamide, 2x SSC (SSC = 0.15M sodium chloride, 0.3M sodium citrate at pH 7), 20 mM Tris at pH 7.5, and 5 mM EDTA in four changes for a total of 2 h at 56°C. After the first wash a digestion step with RNase A (20 μg/ml) and RNase (1 U/ml) (both obtained from Sigma Chemical Co.) for 30 min at 37°C was included. The slides were dipped into NTB-2 nuclear track emulsion (Eastman Kodak, Rochester, NY), 1:2 diluted with 600 mM ammonium acetate, and exposed at 4°C for 8 days. The slides were developed with Kodak developer D-

19 for 2.5 min and fixed with Kodak fixer for 5 min. Counterstaining was done with 4% Giemsa stain (Fisher Scientific Co., Orangeburg, NY) for 10-15 min. From each animal, two sections were each hybridized with a labelled C11 antisense probe (complementary sequence to the cytoplasmic C11 mRNA) and one section was each hybridized with a labelled C11 sense probe.

The results of in situ hybridizations with C11-specific probes demonstrated that the cellular infiltrate in rejecting allografts contains a high proportion of cells expressing C11 transcripts. See Table 1 which shows the frequency of infiltrating cells with detectable levels of C11 mRNA. Cryostat sections of the graft were hybridized with radiolabelled RNA antisense probe of the C11 gene. The results in Table 1 are expressed as the number of positive cells per unit area (1 mm^2) of infiltration area. Three animals with an allograft and two animals with a syngeneic graft were examined and two sections of each animal and each probe were used for evaluation.

The first cells with detectable levels of C11 mRNA were found on day 2 after transplantation both in animals with an allogeneic and those with a syngeneic graft. These positive cells, however, were extremely rare at this timepoint and were normally not found on every section of the same animal. On day 4 after transplantation, the experimental animals showed a 5-10-fold higher frequency of

TABLE 1

Frequency of Infiltrating Cells with Detectable Levels of C11 mRNA

Days after transplantation	Allogeneic graft C11	Syngeneic graft C11
2	3 ± 4	4 ± 1
4	44 ± 69	3 ± 4
6	205 ± 84	7 ± 4
8	313 ± 56	21 ± 12
10	323 ± 112	15 ± 1
12	350 ± 189	3 ± 1

$C11^+$ cells than the control group with a syngeneic graft. The frequency of inflammatory cells expressing the gene increased dramatically between day 4 and 12 after allotransplantation and was at least eight times higher than in the control animals during this period.

In one of the control animals, the syngeneic graft became necrotic and no viable syngeneic graft cells could be detected 8 d after transplantation. This animal, which was not included in Table 1, had 5-10 times more C11 $mRNA^+$ cells than other control animals at the same timepoint. However, compared to the experimental animals 8 days after transplantation, the frequency of positive cells was still ~50% lower. In the first 4 days after the mice received the allograft, about equal numbers of $C11^+$ cells were found among the infiltrating cells.

The amount of C11 specific mRNA per cell, as measured as the number of silver grains over a single cell, increased steadily during the entire observation period in

experimental animals, indicating that the gene was expressed over long periods, perhaps after local induction by alloantigen and/or mediators. In control animals, the expression level increased only slightly after more than 4 days after transplantation.

The phenotype of the C11 transcript positive cells was determined as detailed in Mueller et al. (1988) J. Exp. Med. 167:1124-1136. Briefly, small pieces of the allograft were digested with collagenase, and the resulting suspension of the isolated infiltrating T cells were sorted on a fluorescence activated cell sorter for subsequent in situ hybridization. The infiltrating cells of the allograft and the splenocytes of six animals that received a heart muscle graft 6 days before were pooled and sorted according to their phenotype. The frequency of C11-positive cells in the $CD8^+$ subset was generally 10-20 times higher in the infiltrate of the allograft than in the spleen of the same animals. The recovery of $CD4^+$ cells from the infiltrate was always very low and the frequency of positive cells in this subpopulation was at least 10-fold lower than in the $CD8^+$ subset of infiltrating cells; of 84 C11 $mRNA^+$ cells analyzed, 82 were $CD8^+$ (98%) and 2 $CD4^+$ (2%). On cytospin preparations from sorts cells, C11-transcript positive cells were mainly found among the blast-like $CD8^+$ cells. In double stainings of cell suspensions and tissue sections, no evidence for a significant contribution of $CD4^-$, $CD8^-$ or $CD4^+$, $CD8^+$ T cells among the allograft infiltrating cells and the C11 transcript positive cells were found.

Example 2

Clones B10 and C11 were sequenced according to the dideoxy method of Stanger et al. (1980) J. Mol. Biol. 143:161. Sequence analysis of B10 and C11 (Fig. 2) reveals that they are related to each other and that the

hypothetical proteins they encode contain a short region characteristic of serine proteases, Asp-Ser-Gly-Gly (a sequence homologous to that surrounding Ser^{195} of chymotrypsin).

With B10 and C11 as probes, another CTL complementary DNA (cDNA) library was screened, in which inserts greater than 1000 base pairs were cloned in λgt10. Forty thousand recombinants were screened and 39 plaques corresponding to C11 were isolated.

A cDNA insert of 1400 base pairs, which hybridized with C11, was selected for sequence analysis. The predicted protein sequence encoded, of molecular weight 25,319, is shown in Fig. 3. The putative start codon is preceded by a potential ribosome binding site CCUUCCG (Hagenbuchle et al. (1978) Cell 13:551) and a polyadenylation signal sequence AAUAAA (Proudfoot & Brownlee, (1966) Nature 263:211) occurs just upstream from the poly(A) tract. Of the first 12 amino acids predicted, ten are hydrophobic, and the amino acid in position 2 (Lys) is basic, suggesting that this sequence may act as a signal to direct secretion or intracellular organelle location. A search of the National Biomedical Research Foundation (NBRF) protein sequence data bank revealed that the protein encoded by C11 resembles a number of serine proteases (Table 2).

Table 2

Protein	E.C. number	Species source	Residues compared		Percent homology
			CCP1	Bank protein	
7S nerve growth factor	3.4.21	Murine	29-224	26-229	40
Chymotrypsin A	3.4.21.1	Bovine	1-200	16-216	35
Chymotrypsin B	3.4.21.1	Bovine	1-200	16-216	36
Complement C1r	3.4.21.41	Human	52-224	56-238	35
Elastase	3.4.21.11	Porcine	3-220	3-233	33
Factor X	3.4.21.6	Bovine	1-225	192-421	33
RMCPII	3.4.21	Rat	1-214	1-213	51
Kallikrein	3.4.21.8	Rat	26-225	51-262	36
Plasminogen	3.4.21.7	Human	3-224	563-787	37
Plasminogen activator	3.4.21.31	Human	72-224	389-560	35
Trypsin	3.4.21.4	*S. griseus*	29-220	22-214	33
Trypsin	3.4.21.4	Rat	29-226	31-228	39

When the sequences were optimally aligned according to the Dayhoff algorithm (Dayhoff, in Atlas of Protein Sequencing and Structure 5:1 (Supp. 3) (National Biomedical Res. Found., Washington, D.C., 1979)), the homologies generally varied between 30 and 40 percent. The greatest homology was found with rat mast cell protease type II (RMCPII), which had amino acids identical to 109 of 215 amino acids encoded by C11, giving a match per length of 51 percent. The amino acid residues known to form the catalytic triad of the active site in serine proteases (His^{57}, Asp^{102}, and Ser^{195}) were all found in the protein encoded by C11 (Fig. 3, A). The sequences around these residues, which are highly conserved among serine proteases, are also conserved in the C11 gene product. Indeed, largely because of conservation around this region, the protein encoded by C11 appears to be somewhat homologous (about 30 percent of 209 residues) even to the prokaryotic proteases trypsin and type B from Streptomyces griseus.

The cytotoxic T lymphocyte-specific proteins (CCP's) encoded by C11 and B10 will be referred to as CCP1 and CCP2, respectively. In Fig. 4 the optimal protein alignment with CCP1 is presented for RMCPII, bovine chymotrypsin, bovine trypsin, and CCP2 (not numbered, as the full sequence is not presented). The full sequence of CCP2 can be obtained by application of the procedures applied to C11 and CCP1. The full sequence of CCP2 is presented in Fig. 5.

RMCPII is an intracellular serine protease found in the granules of atypical mast cells. The high level of homology of CCP1 with RMCPII is particularly intriguing as RMCPII has a number of structural features that make it exceptional in the serine protease superfamily. Protein CCP1 contains cysteines in precisely the same positions as RMCPII which, by analogy with RMCPII, form three disulfide

bonds. These occur in the same positions in chymotrypsin, trypsin, and elastase. Both CCP1 and RMCPII lack a disulfide bond that is present in all other known serine proteases, including several from prokaryotes, and that links Cys^{191} with Cys^{220} in chymotrypsin. In both CCP1 and RMCPII the first of these two half-cysteines is replaced by a phenylalanine, while the second half-cysteine has been deleted along with other residues. Linkage of Cys^{191} to Cys^{220} is thought to be important in stabilizing the conformation of the substrate binding site (Woodbury et al., (1978) Biochem. 17:811). Its absence in CCP1 and RMCPII may lead to significant changes in that site and, hence, in substrate specificity.

Two other primary structure changes previously seen only in RMCPII and thought to alter substrate binding are also present in the predicted CCP1 protein. In RMCPII and CCP1 the amino acid six residues before the active-site serine is alanine. In chymotrypsin-like proteases it is serine and in trypsin-like proteases, aspartic acid. The residue in this position lies at the bottom of the S_1 binding site, so the change to a less polar residue would indicate a preference for a hydrophobic amino acid at the P_1 position in the substrate. Furthermore, the sequence $Ser-Trp-Gly^{216}$ in chymotrypsin, which forms hydrogen bonds with the P_1 and P_3 residues of the substrate, is replaced by Ser-Tyr-Gly in CCP1 and RMCPII, again suggesting altered substrate specificity. Both of these changes are also seen with CCP2.

One of the few RMCPII-specific differences that is not present in CCP1 is the substitution of isoleucine at position 99 in chymotrypsin for asparagine. In most mammalian serine proteases this residue is hydrophobic, and indeed in CCP1 it appears to be phenylalanine. However,

most of the RMCPII-specific changes are present in CCP1 protein, suggesting that the substrate binding site of CCP1 resembles that of RMCPII and is significantly different from those of other mammalian serine proteases.

Example 3

The initial step in determining the structure of a protease expressed exclusively by human cytotoxic T lymphocytes and recognizing a unique protein cleavage site is to clone human cytotoxic T lymphocyte specific cDNAs.

PolyA$^+$ RNA from a human cytotoxic T lymphocyte cell line, e.g., one of the lines on deposit at the Coriel Institute for Medical Research, Copewood and Davis Street, Camden, NJ, is used as a template for the synthesis, by standard procedures, of double stranded complementary DNA. EcoRI recognition sequences are then ligated onto the ends of the dscDNA by standard methods, and the resultant molecules are size selected on low melt agarose and then inserted into the EcoRI site of λgt11, all by conventional procedures. These recombinant molecules are then packaged into λ phage heads (Gigapack plus, Stragene) and used to infect E. coli Y1088. DNA from plaques harboring recombinant molecules are hybridized with radioactive probes generated from B10 and C11 by standard procedures to identify corresponding human genes. The screening is conducted in duplicate to minimize the possibility of false positives. hC11, a human counterpart of C11, was found using the above procedures.

The phage DNA from any positive plaques are isolated and immediately recloned, using conventional procedures, in the plasmid vector pUC13. Large amounts of these recombinant plasmid DNAs are then isolated for further analysis. The human cytotoxic T lymphocyte specific clones can be characterized by restriction enzyme digestions and,

ultimately, sequence analysis. In addition, their relationships to one another can be investigated by standard cross-hybridization and heteroduplex mapping.

Tissue specific expression and transcript sizes of isolated genes can be established using the same methods as described for B10 and C11. Using Northern blot analysis, as described above, a number of different cell lines (all obtained from ATCC) were tested for expression of hC11. CEM-CM3 (acute lymphoblastic leukemia), CCRF-CEM (acute lymphoblastic leukemia), CCRF-SB (acute lymphoblastic leukemia), RPMI 7666 (B lymphoblast), DLD-1 (colon adenocarcinoma), and CRL-7123 (spleen line) all failed to express hC11. Human thymocytes and peripheral blood lymphocytes were also negative. Cytolytic T cells, activated by mitogen, interleukin 2, anti-T cell-receptor antibody, or fucose, were all positive for hC11 expression. A human cytotoxic T cell line was also positive. Thus, expression of hC11 appears to be specific to cytotoxic T cells.

The correlation between the level of cytotoxicity and the expression of the human genes also can be examined using the above-described methods. The expression of hC11 was found to correlate with the cytolytic activity of the cells in which it was expressed. Expression of hC11 was detected in lymphokine activated killer (LAK) cells. The procedure for generating LAK cells is essentially that of Rosenberg et al. (1985) N.E.d. Med 313:1485.

When expression of a human cytotoxic T lymphocyte-specific gene correlates with toxicity, the gene is sequenced by standard methods (as was done with the B10 and C11 genes). From the gene sequence, the structure of the protease can be determined, and a computer analysis of the structure of protease performed, as with the C11 gene.

Further computer analysis can show the location of the active site of the enzyme, and the appropriate sequence of a peptide that can act as a competitive inhibitor can be determined.

hC11 was sequenced, as described above, and found to be very similar to murine gene C11 (Fig. 6). The active site of hCCP1, the protein encoded by hC11, resembles the active site of the murine protein, CCP1, very closely. Most importantly, like CCP1, hCCP1 appears to have an Arg at S_1, imparting the unusual specificity of Asp at P_1. The only other difference is the substitution of an aromatic amino acid two residues downstream from the Arg. Due to the similarity of the proteins encoded by hC11 and C11 inhibitors synthesized to inhibit one should inhibit the other.

A partially purified preparation of hCCP1 does not cleave at sites recognized by trypsin and chymotrypsin.

Analysis of hC11 gene expression, by in situ hybridization to biopsy sample, indicates that hC11 is expressed in cardiac tissue of a patient that rejected a transplanted heart. In situ hybridization and related procedures were performed as described above.

Example 4

A human placental genomic library, in λ charon 4A, was screened by hybridization in 20% formamide and 6 x SSC (1 x SSC is 0.15 M sodium chloride, 0.3 M sodium citrate, pH7) at 41°C with a mixture of radioactivity labelled cDNAs corresponding to the murine cytotoxic cell proteases CCP1-4, Bleackley et al., (1988) FEBS Letters 234: 153-159 and Lobe et al. (1976) Science 232: 858-861.

Phage DNA from one of the positive plaques gave a 6.3 kb EcoRI fragment (and ultimately a 1.5 kb Bam fragment) (Fig. 7) that hybridized with the murine probes but failed,

under conditions of high stringency, to hybridize with hCCP1. Preliminary sequence analysis revealed that the 1.5 kb fragment encoded a protein which was highly homologous to the murine cytotoxic proteases. Thus this gene is a new member of the human CCP family but is different from hCCP1.

hCCPX is expressed in cytotoxic cells. Poly A^+ RNA was purified from resting and activated peripheral blood lymphocytes and subjected to Northern blot analysis using the 1.5 kb genomic fragment as a probe. A transcript is clearly present in the activated cells that is absent in RNA from the unstimulated control. Sometimes a small amount of transcript is seen in the unstimulated cells, perhaps due to cellular contamination, however, the transcript is always induced upon stimulation.

Because of the high level of homology between the various CCP family members cross-hybridization can occur. In the case of the murine genes, CCP1 can be distinguished from the others because of a difference in transcript size. However, the transcripts detected by hCCP1 and HCCPX are very similar in mobility. Therefore, high stringency washing conditions were used to minimize cross-hybridization. With washing at 41°C the 1.5 kb probe detects transcripts in both human and mouse cytotoxic cells. However at 55° the signal due to the cross-hybridization with the mouse transcripts is markedly less than that seen for the human RNA, even though this mouse cell line expresses extremely high levels of the protease transcripts. The human-human and human-mouse identities are both approximately 70%, thus we believe that the signal seen under high stringency washing conditions in the RNA from activated human cells is due to specific hybridization with hCCPX transcripts.

In addition, no detectable signal was detected using this probe on RNA samples from a number of human cell lines obtained from the ATCC including CEM-CM3, CCRF-CAM, CCRF-SB (acute lymphoblastic leukemias), RPMI 7666 (EBV-transformed B lymphoblast), DLD-1 (colon, adenocarcinoma), CRL-7020 (thymus), CRL-7123 (spleen) and freshly isolated human splenocytes and thymocytes.

Example 5

The nucleotide sequence of the region indicated by the heavy line in Fig. 7 is presented in Fig. 8. A comparison of this sequence with those of the murine CCP genes revealed high levels of homology (~70% identity) in regions which correspond to exons and dissimilarity in regions which correspond to introns. By placing the introns (the underlined regions in Fig. 8) in exactly the same places that they occur in the murine sequences (all four murine genes have introns in precisely the same positions, Lobe et al. (1988) Biochemistry 27: 6941-6946), the sequence of a cDNA could be determined (Fig. 9). A cDNA corresponding to exons 3, 4 and 5 has been isolated and confirms the positioning of the introns. The predicted protein which would be encoded by this gene is 246 amino acids in length (molecular weight = 27,318). The amino acid sequence is shown below the nucleotide sequence in Fig. 9. This protein was not found in the GenBank data base. It is however, homologous to a wide variety of serine proteases. The highest level of identity was with the cytotoxic cells proteases (human 70%), murine (61%), cathepsin G (human 57%), and mast cell proteases (40-50%). In addition, a significant level of identity (~30%) was found with many other trypsin and chymotrypsin like enzymes. This protein is a serine protease and is related to the cytotoxic cell

proteins, it will be referred to as human cytotoxic cell protease-X (hCCPX).

An alignment of the hCCPX sequence with those predicted from the murine genes (Fig. 10) illustrates the high degree of primary sequence similarity and also reveals that hCCPX shares many features in common with the CCP genes, Bleackley et al. (1988) FEBS Letters 234: 153-159. hCCPX is very basic (14% basic, 6% acidic amino acids) and contains a hydrophobic leader sequence of 18 residues followed by a putative zymogen dipeptide which precedes the mature protease amino terminal Ile residue. It is believed that the basic nature of the proteins may play a role in sequestering them within granules bound to proteoglycans, Stevens et al. (1988) Current Topics in Microbiology and Immunology 140: 93-108. The two sequences +21 to +24 (Ile Ile Gly Gly) and +29 to +36 (Pro His Ser Arg Pro Tyr Met Ala) which are found in all the CCPs, granzymes, RMCPI and II, and cathepsin G are also conserved in hCCPX as are the six cysteine residues which form disulfide bonds, Jenne et al., (1988) Current Topics in Microbiology and Immunology 140:33-48. The catalytic triad residues (marked with an "*" in Fig. 10) which form the active site of the serine proteases are all present in the correct positions, Neurath (1984) Science 224:350. The sequences surrounding these, which are highly conserved in serine proteases, are also conserved.

CCP1 and 2 both contain unusual residues in regions that are believed to be important in defining substrate specificity, Lobe et al. (1986) Science 232: 858-861 and Murphy et al. (1988) Proteins 4:190-204. In addition, they lack a disulfide bond which in other serine proteases is important in restricting the size of the substrate binding pocket. Similar results were subsequently found for the

other CCPs and granzymes, Bleackley et al. (1988) FEBS Letters 234:153-159 and Masson et al. (1987) Cell 49:679-685. The protease described here also has unusual residues in these same sites and lacks the disulfide bond. However, the pattern of amino acids seen in this protein, namely Thr, Ser-Tyr-Gly, and Gly at positions -6, +15 to +17, and +25 relative to the active site Ser, does not correspond to any of the murine proteases characterized to date. It would appear then that hCCPX would also have an unusual substrate specificity.

Purified insert from the cDNA containing plasmid was labelled by random priming and used as a probe for in situ hybridization on human metaphase spreads. The gene is present at a single locus on chromosome 14 at q11.2. The human gene encoding hCCP1 maps to the same region. In mice the genes encoding CCP1, CCP2, CCP3, and CCP4 are all located on chromosome 14 close to the α-chain of the T cell antigen receptor locus Brunet et al. (1986) Nature 322:268-271.

Example 6

The three-dimensional structure of CCP1, the protease encoded by C11, was predicted by computer analysis. The use of comparative molecular modeling to predict the structure of a protease and its characteristic substrate is particularly reliable when the protein of unknown structure is relatively homologous with a protein of known three dimensional structure. The existence of a large database of known three dimensional structures of related proteins and their substrate is also very helpful. In the case of the cytotoxic cell proteases both of these criteria are met.

The model building procedure (as applied to CCP1 and another unrelated serine protease) is described in detail in

Murphy et al. (1988) Proteins: Structure, Function, and Genetics 4:190-204 which is included herein as an appendix. (The computer program MUTATE referred to in the appendix is available from Dr. R. Read, Department of Medical Microbiology, University of Alberta, Canada). Briefly, the process begins with aligning the sequence of the protein of unknown structure with the sequence of a template protein, a protein of known three-dimensional structure. In the case of highly homologous proteins the alignment is straightforward: the sequences are aligned and a computer generated model of the template protein is modified to yield a model of the structure of the unknown protein. The side chain of each amino acid of the template is then replaced with the side chain of the corresponding amino acid of the protein of unknown structure. The replacement side chain conformations are adjusted to follow the conformation of the replaced, i.e., template, side chain conformations when possible. When this is not possible preferred side chain angles are selected from a dictionary of preferred side chain conformations.

Subsequent refinements include adjusting the model to remove unacceptably close non-bonded intramolecular contacts and adjusting the placement of deletion and insertion loops. In the final step, the deduced structure is adjusted to relieve any remaining unacceptably close non-bonded contents.

The prediction of substrate structure is drawn from several types of information. This procedure begins with an examination of the deduced three dimensional structure of the protease and an analysis of the identity of amino acid residues in key positions on the catalytic site of the protease. This information is compared to the reactive site on the substrate of a closely related protease. The

sequence of that substrate can then be altered to achieve a sequence complementary to the catalytic site of the modeled protein.

Analysis of CCP1 indicates that the active site has histidine, aspartic acid, and serine residues at the back, meaning that it is a serine protease. Computer analysis further indicated that this active site cleaves proteins at a cleavage site (between the C-linkage of an Asp residue and an N-linkage of an adjacent amino acid, Phe) different from the cleavage sites recognized by any other known eukaryotic serine proteases. This deduced cleavage site permits the synthesis of synthetic peptides which, by mimicking all or a portion of the natural cleavage site, can bind to the active site of the protease and competitively inhibit it.

Example 7

The amino acid residues of a substrate are designated $P_4P_3P_2P_1P_1'P_2'P_3'P_4'$ with cleavage by the protease occurring between P_1 and P_1'. The corresponding interacting amino acids of the binding pocket of an enzyme are designated $S_4S_3S_2S_1S_1'S_2'S_3'S_4'$ with S_1 for e.g., example interacting with P_1.

The computer generated three-dimensional structure of CCP1 indicates that the residues of the binding pocket which might interact with a substrate are: Pro 28-Cys 42; His 57-Asn 65; Leu 32; Ile 41; Ile 73; Tyr 151; Gly 153; Phe 99; Ser 214-Asp 219; Phe 191-Ser 195; Arg 226; and Asn 174-Arg 175. (See pages 198-200 of Murphy et al., Appendix). The most important prediction is that S_1 equals Arg 226. This predicts an acid substrate specificity (probably Asp) at P_1, the site of cleavage. This specificity is unique among eukaryotic serine proteases. S_2 appears from the computer analysis to be Phe 99, indicating a small amino acid e.g., Val, at P_2. The presence of basic residues in S_3

and S_4 predict acidic residues at P_3 and P_4. Guided by these considerations the inhibiting peptides, corresponding to residues P_3-P_1' of the substrate, were synthesized. These peptides are shown in Fig. 11. The effect of the inhibitors on the cytotoxic properties of cytotoxic T lymphocytes is shown in Tables 3 and 4.

TABLE 3

	% Lysis			
	2368	2369	2372	2373
Control cytotoxicity	15%	15%	15%	15%
+ 100 μg/ml peptide	8%	13%	12%	9%
Control cytotoxicity	30%	30%	30%	30%
+ 50 μg/ml peptide	19%	17%	17%	23%

Cytotoxicity was measured with cells from a cyclosporine-A induced mixed lymphocyte reaction mixed lymphocyte reaction (CsA-MLR). Spleen cells were obtained aseptically by pressing the spleen through a wire mesh into a medium of RPMI 1640 (GIBCO Laboratories, Grand Island, NY), 10% (v/v) fetal bovine serum (GIBCO Laboratories), 10^{-4}M 2-mercaptoethanol, and 10 mM HEPES buffer (Sigma, St. Louis, MO) (RHFM). Responder cells ($1-2x10^6$/ml) were cocultured with equal numbers of allogenic stimulator cells (1500 rad from a ^{137}Cs source) in RHFM plus 300mg/ml CsA and 200 units/ml interleukin 2 in a final volume of 4 ml (Costar 6 well cluster) or 25 ml (Costar 75 cm^2 tissue culture flask). The cultures were incubated at 37°C in 5% CO_2 and 90% relative humidity. Cells from the primary MLR cultures were harvested, washed in RHFM and then recultured with cytokines at a cell density of $2-5 \times 10^5$ cells/ml for 24 or 48 hours. In some experiments, viable cells were isolated by gradient density centrifugation. For cytotoxicity assays, cells were incubated with 10^4 target cells labelled with $Na^{51}CrOI_4$ (New England Nuclear, Boston, MA) in a round-bottom microtiter plate (final volume of 200 μL). After 4 hours at 37°, 100 μL of supernatant was removed from each well for counting. Specific lysis was calculated as:

$$\% \text{ lysis} = \frac{\text{experimental - spontaneous release}}{\text{total release - spontaneous release}} \times 100$$

Spontaneous release was obtained by incubating ^{51}Cr-labelled targets alone, and total release from target cells incubated with 1% Zap-Isoton lytic agent (Coulter Electronics of Canada, Ltd., Mississauga, Ontario).

TABLE 4

Peptide	Control	Untreated effectors	Pretreated effectors
EF2394	54%	90%	58%
EF2395	61%	86%	78%
EF2396	54%	73%	50%
EF2397	54%	87%	50%
EF2398	54%	107%	89%
EF2368	54%	100%	75%
EF2369	54%	100%	28%
EF2372	54%	97%	35%
EF2373	54%	87%	40%

Cytotoxicity in cytotoxic T-cells activated with ConA and interleukin 2 (IL2). Cytotoxic T Cells were activated using 10 µg/ml ConA and 10 U/ml IL2. The cytolytic activity was measured in a standard chromium release assay. Targets were pretreated for 3 hr with 100 µg/ml peptide and were then mixed with either pretreated 100 µg/ml peptide and were then mixed with either pretreated 100 µg/ml peptide) or untreated effectors. All results are at an effector to target ratio of 5:1. Results are calculated as described in the legend of Table 1.

Example 8

The cDNA clones of the invention can be used to generate copious quantities of purified cytotoxic cell proteases by inserting the coding sequence of a cytotoxic cell protease gene into an expression vector and expressing the desired proteins in an expression system. These procedures are well known to those skilled in the art.

The possession of purified protease allows for a greatly simplified alternative approach to the design of inhibitor molecules. Rather than the extremely cumbersome and complex immunologically based assays used to produce the results in Tables 3 and 4, the enzymatic action of the purified protease on a given substrate can be followed directly, by cleavage of the substrate, when the purified protease is available. (Sequence specific protease cleavage can be followed with standard thioester-based assays such as that described in Harper et al. (1984) Biochem. 23:2995-3002). This allows a large number of potential inhibitors to be tested with relative ease. The purified protease based assay can be used alone, or in conjunction with the rational design factors obtained by computer analysis, to screen large numbers of potential inhibitors. Positive compounds could then be tested for their immunosuppressive properties.

Inhibitor Peptide Synthesis

The inhibitory peptides of the invention can be prepared by standard solid phase synthesis, for example, a method in which a tert-butyloxycarbonylamino acid is attached to either chloromethyl resin containing 0.75 mM Cl^- g^{-1}, or the p-methylbenzhydrylamine resin containing 0.35 mM NH_2 g^{-1}, followed by the sequential addition of desired amino acid residues to produce the desired peptide. Synthetic reactions are performed in 70 ml polypropylene syringes fitted with a polyethylene frit using apparatus and techniques described in Burton et al., (1975) Biochemistry 14:3892, and Merrifield, (1963) J. Amer. Chem. Soc. 85:2149. Completeness of coupling is determined by the standard ninhydrin test. The C-terminal amino acid is attached using procedures described in Stewart et al., Solid Phase Peptide Synthesis (W.H. Freeman ed. 1970), or Pietta et al., 1970

Chem. Comm. 650. Hplc purifications of the synthetic peptides are carried out using a Beckman ODS column (10 x 250 mm).

Amino acid analyses of the synthetic peptides are, if desired, performed using a Durrum D-500 analyzer. Cysteinyl residues in the peptides are quantitated as cysteic acid using a modification of the method of Moore (1968) in which 100 mM peptide is oxidized with 2.0 ml performic acid (1 ml 30% H_2O_2 + 9 ml 88% HCOOH) for 2 hrs. at 0°. Performic acid is removed in a Reacti-Therm at 40° using N_2, and 0.5 ml distilled water is then added to the residue and re-evaporated. The product is then hydrolyzed using 6 N HCl. Free sulfhydryl groups are determined using the method of Ellman et al. (1959).

Use

The inhibitory molecules are effective inhibitors of cytotoxic cells, e.g., cytotoxic lymphocytes. The inhibition of the target cell destroying activity of such cells can be used to treat patients suffering of autoimmune diseases such as Hashimoto's thyroiditis, primary myxedema, thyrotoxicosis, pernicious anaemia, autoimmune atrophic gastritis, Addison's disease, myasthenia gravis, juvenile diabetes, Goodpasture's syndrome, pemphigus vulgaris, pemphigoid, sympathetic ophthalmia, phacogenic uveitis, autoimmune haemolytic anaemia, idiopathic thrombocytopenic purpura, idiopathic leucopenia, primary biliary cirrhosis, active chronic hepatitis HB_S-ve, cryptogenic cirrhosis (some cases), ulcerative colitis, Sjögren's syndrome, systemic lupus erythematosus (SLE), discoid LE, dermatomyositis, scleroderma, rheumatoid arthritis, and possibly multiple sclerosis, and similar diseases in other mammals, for example, various types of livestock such as cows. Such inhibition can also be used to treat allograft (a tissue or

organ graft from a donor who is a genetically dissimilar member of the same species as the receptor) rejection, and graft v. host disease.

The peptides can be administered to a mammal in a dosage of 25 to 500 mg/kg/day, preferably 50 to 100 mg/kg/day. When administered to mammals (e.g., orally, intravenously, parenterally, nasally, or by suppository), the peptides inhibit the ability of cytotoxic T lymphocytes to destroy cells, thus inhibiting the cell-mediated immune response to provide an effective treatment for the above listed disorders.

Nucleic acid probes (prepared by standard methods) capable of hybridizing to a gene encoding a protease expressed only by cytotoxic lymphocytes can be used in a variety of useful hybridization assays. For example, such probes can be used to monitor cytotoxic T lymphocytes in transplanted tissue, e.g., by the in situ hybridization methods of Cox et al. (1984) Dev. Biol. 101:485. The presence of the lymphocytes in the transplanted tissue is an indication that the tissue is being rejected by the host organism and that appropriate immunotherapy should be undertaken.

The probes can also be used to assess the potential cytotoxicity of lymphokine activated killer cells. The generation and use of such cells to treat tumor patients is described by Rosenberg et al. (1985) N.E.J. Med. 313:1485. Rosenberg describe how human peripheral-blood lymphocytes are treated with interleukin-2 (a lymphokine) to generate killer cells that will attack tumor cells when reintroduced into the host. The probes can be used in a hybridization assay with the nucleic acid of the treated lymphocytes by standard methods; the assay monitors the degree to which the activated killer cells have been generated by

determining the level of expression of the protease-encoding gene in the cells.

Other embodiments are within the following claims.

What is claimed is:

Claims

1. A vector comprising a DNA sequence encoding the CCP1 protein.

2. A vector comprising a DNA sequence encoding the CCP2 protein.

3. A vector comprising a DNA sequence encoding the hCCP1 protein.

4. A vector comprising a DNA sequence encoding the hCCPX protein.

5. A substantially pure CCP1 protein expressed from the vector of claim 2.

6. A substantially pure CCP2 protein expressed from the vector of claim 3.

7. A substantially pure hCCP1 protein expressed from the vector of claim 4.

8. A substantially pure hCCPX protein expressed from the vector of claim 5.

9. A peptide of the formula:
Asp-Val-Asp-Ala;
Ala-Pro-Asp-Ala;
Ala-Asn-Pro-Ala;
Phe-Pro-Arg-Phe;
Ala-Pro-Arg-Phe;
Phe-Pro-Asp-Phe;

Phe-Pro-Asn-Phe;

Phe-Asn-Pro-Phe; or

Phe-Asp-Pro-Phe.

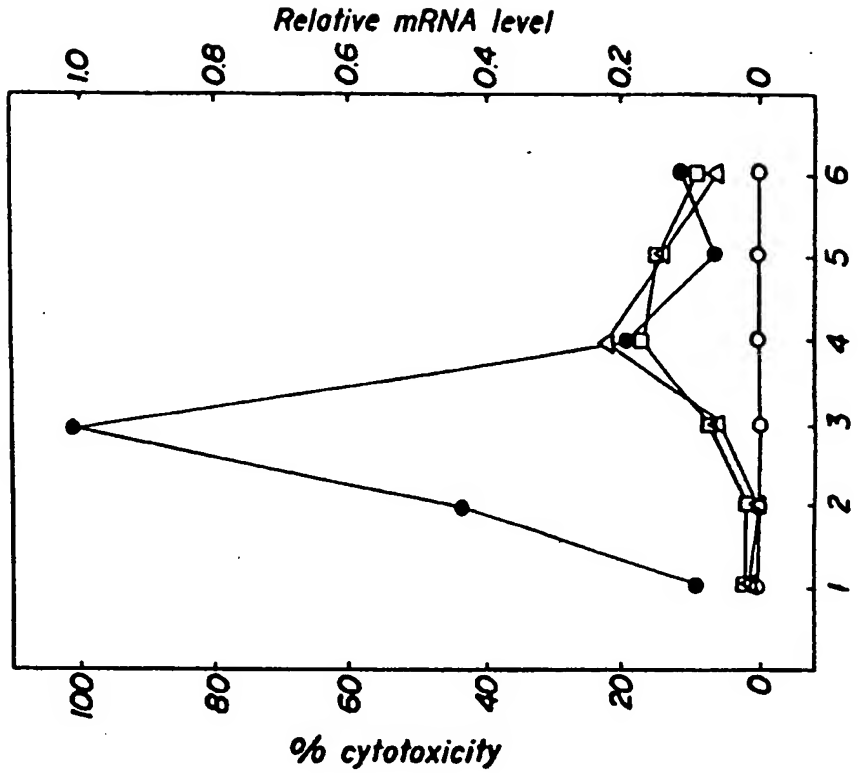

Relative mRNA level
1.0
0.8
0.6
0.4
0.2
0
Days after stimulation
1
2
3
4
5
6
% cytotoxicity
100
80
60
40
20
0

FIG. 1B

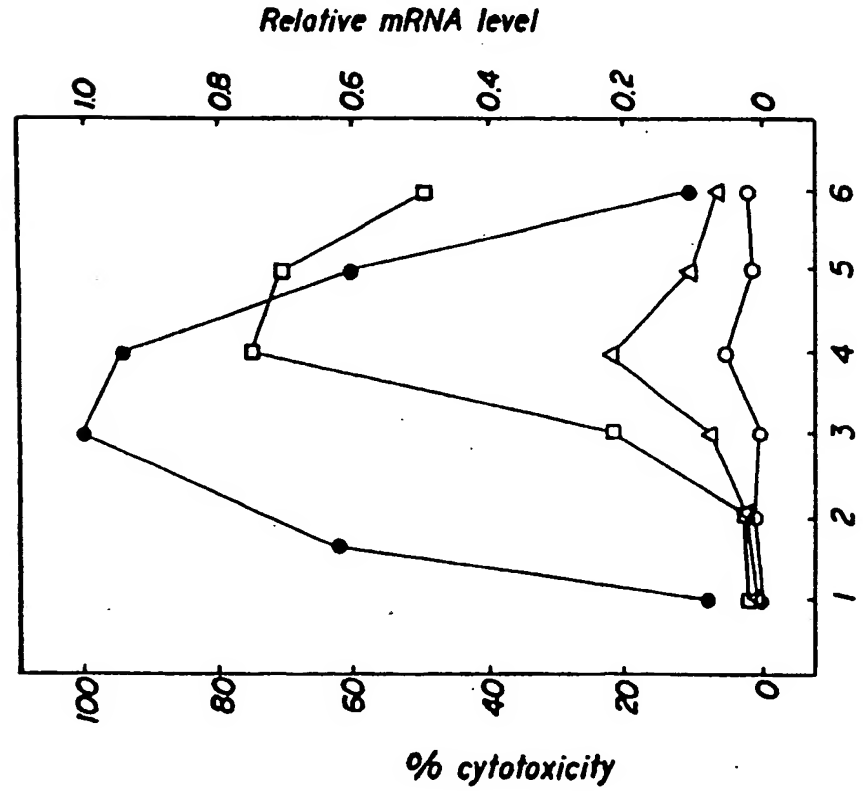

Relative mRNA level
1.0
0.8
0.6
0.4
0.2
0
Days after stimulation
1
2
3
4
5
6
% cytotoxicity
100
80
60
40
20
0

FIG. 1A

```
B10
1
CAAAAACACTGCACGAAGTTAAGCTGACAGTACAGAAGGATCAGGTGTGTGAGTCCCAGTTCCAAAGTTTTTACAACAGAGCTAATGAGATATGTGTGGG

C11
485
CAAACACGCTACAAGAGGTTGAGCTGACAGTACAGAAGGATCGGGAGTGTGAGTCCTACTTTAAAAATCGTTACAACAAAACCAATCAGATATGTGCGGG
    *  *  *     *   *                      *  *          * *  **   * **        * * *   *         *

B10
101
AGACTCAAAGATCAAGGGAGCTTCCTTTGAGGAGGATTCTGGAGGCCCGCTTGTGTGTAAAAGAGCAGCTGCAGGCATCGTCTCCTACGGGCAAACTGAT

C11
585
GGACCCAAAGACCAAACGTGCTTCCTTTCGGGGGGATTCTGGAGGCCCGCTTGTGTGTAAAAAAGTGGCTGCAGGCATAGTTTCCTATGGATATAAGGAT
*   *      *   ** *         **  * AspSerGlyGly                 *  **          * *    *  ** * **

B10
201
GGATCAGCTCCGCAAGTCTTCACAAGAGTTTTGAGTTTTGTATCGTGGATAAAGAAAACGATGAAACACAGCTAACTACAAGAAGCAAC TAGATCCTG

C11
685
GGTTCACCTCCACGTGCTTTCACCAAAGTCTCGAGTTTCTTATCCTGGATAAAGAAAACAATGAAAAGCAGCTAACTAC AGAAGCAACATGGATCCTGC
  *   *    * ** **    * *   * *     **   *              *      **          *         * *     *

B10
299
ACTGA CAGCCATCTTCCC ATAGCTGAGTCCAGGATTGCTCTAGGACAGATGGCAGGCAACTGAATAAAGAACTTTCTCTGACTGCAAAAAAAAAA 393

C11
784
TCTGATTACCCATCGTCCCTAGAGCTGAGTCCAGGATTGCTCTAGGACAGGTGGCAGG ATCTGAATAAAGGAC TGCAAAGACTGGCTTCATGTCC 878
*    ** *     *    * *                        *       * *         *  * * ***     ***** *****
```

FIG. 2


Sau3a
50
Sau3a
100
GGCCITCCGGGGAAGATGAAGATCCTCCTGCTACTGCTGACCTTGTCTCTGGCCTCCAGGACAAAGGCAGGGGAGATCATCGGGGGACATGAAGTCAAGCCC
MetLysIleLeuLeuLeuLeuThrLeuSerLeuAlaSerArgThrLysAlaGlyGluIleIleGlyGlyHisGluValLysPro
1
Sau3a Sau3a
150
Sau3a
200
CACTCTCGACCCTACATGGCCTTACTTTCGATCAAGGATCAGCAGCCTGAGGCGATATGTGGGGGCTTCCTTATTCGAGAGGACTTTGTGCTGACTGCTGCT
HisSerArgProTyrMetAlaLeuLeuSerIleLysAspGlnGlnProGluAlaIleCysGlyGlyPheLeuIleArgGluAspPheValLeuThrAlaAla
20
40
250
300
CACTGTGAAGGAAGTATAATAAATGTCACTTTGGGGGCCCACAACATCAAAGAACAGGAGAAGACCCAGCAAGTCATCCCTATGGTAAAATCCATTCCCCAC
HisCysGluGlySerIleIleAsnValThrLeuGlyAlaHisAsnIleLysGluGlnGluLysThrGlnGlnValIleProMetValLysCysIleProHis
60
350
400
CCAGACTATAATCCTAAGACATTCTCCAATGACATCATGCTGCTAAAGCTCAAGAGTAAGGCCAAGAGGACTAGAGCTGTGAGGCCCCTCAACCTGCCCAGG
ProAspTyrAsnProLysThrPheSerAsnAspIleMetLeuLeuLysLeuLysSerLysAlaLysArgThrArgAlaValArgProLeuAsnLeuProArg
80
100
450
500
CGCAATGTCAATGTGAAGCCAGGAGATGTGTGCTATGTGGCTGGTTGGGGAAGGATGGCCCCAATCGGCAAATACTCAAACACGCTACAAGAGGTTGAGCTG
ArgAsnValAsnValLysProGlyAspValCysTyrValAlaGlyTrpGlyArgMetAlaProMetGlyLysTyrSerAsnThrLeuGlnGluValGluLeu
120
140
Sau3a
550
600
ACAGTACAGAAGGATCGGGAGTGTGAGTCCTACTTTAAAAATCGTTACAACAAAACCAATCAGATATGTGCGGGGGACCCAAAGACCAAACGTGCTTCCTTT
ThrValGlnLysAspArgGluCysGluSerTyrPheLysAsnArgTyrAspLysThrAsnGlnIleCysAlaGlyAspProLysThrLysArgAlaSerPhe
160
650
700
CGGGGGGATTCTGGAGGCCCGCTTGTGTGTAAAAAAGTCGCTGCAGGCATAGTTTCCTATGGATATAAGGATGGTTCACCTCCACGTGCTTTCACCAAAGTC
ArgGlyAspSerGlyGlyProLeuValCysLysLysValAlaAlaGlyIleValSerTyrGlyTyrLysAscGlySerProProArgAlaPheThrLysVal
180
200

FIG. 3A

```
                                750                                   800
TCGAGTTTCTTATCCTGGATAAAGAAAACAATGAAAAGCAGCTAACTACAGAAGCAACATGGATCCTGCTCTGATTACCCATCGTCCCTAGAGCTGAGTCCA
SerSerPheLeuSerTrpIleLysLysThrMetLysSerSer***

                           850                                   900
GGATTGCTCTAGGACAGGTGGCAGGATCTCAATAAAGGACTGCAAAGACTGGCTTCATGTCCATTCACAAGGACCAGCTCTGTCCTTGGCAGGCCAATGGAA

                      950                                   1000
CACCTCTTCTGCCACCATGCTGTGACAACCCAACTGACATCTTCCTATGGAAGTTTGCCCTCTCCACAAAAGAAGTAGAATGTTTGCATTGGAGCTGGGCAT

                 1050                                  1100
GCTCTGCTTCCCCTCAGTGCCCCGAGAATGTTATCTAATGCTAGTCATCATTAATAGCTCCCTACAGAACTTTCATACAGTTGCACCCAAGTTGCTGATGTG

            1150   EcoRI                          1200
TTCTCTAGAATAGAGCAAGAAATAGTAAACAGAATTCCTTTTGCCTCTCTGTACTATTTTCCCCCAAATACCAAGATTTGTATGTTTTATAAAGCTAATTTC

       1250                                  1300   Sau3a
CTTATCAAATGACATCTTTTAATTTTTACATTAATGGCTTATTTTCAAGGTACAACCTGATTTTTTTATGGACAAAAATGATCGTAAAATCAAATAAAACTA

  1350
ATTAATATATCCAAAAAAAAAAAAAAAA
```

FIG. 3B

```
CCPI    Ile Ile Gly Gly His Glu Val Lys Pro His Ser Arg Pro Tyr Met Ala Leu Leu Ser Ile Lys Asp  22
RMCPII  Ile Ile Gly Gly Val Glu Ser Ile Pro His Ser Arg Pro Tyr Met Ala His Leu Asp Ile Val Thr  22
CA COW  Ile Val Asn Gly Glu Glu Ala Val Pro Gly Ser Trp Pro Trp    Gln Val    Ser Leu Gln Asp  37
TR COW  Ile Val Gly Gly Tyr Thr Cys Gly Ala Asn Thr Val Pro Tyr    Gln Val    Ser Leu Asn Ser  26

CCPI    Gln Gln Pro Glu    Ala Ile Cys Gly Gly Phe Leu Ile    Arg Glu Asp Phe Val Leu Thr Ala  42
RMCPII  Glu Lys Gly Leu Arg Val Ile Cys Gly Gly Phe Leu Ile Ser Arg Gln    Phe Val Leu Thr Ala  43
CA COW  Lys Thr Gly Phe    His Phe Cys Gly Gly Ser Leu Ile    Asn Glu Asn Trp Val Val Thr Ala  55
RW COW  Gly Tyr            His Phe Cys Gly Gly Ser Leu Ile    Asn Ser Gln Trp Val Val Ser Ala  44

CCPI    Ala His Cys    Glu Gly Ser Ile Ile Asn Val Thr Leu    Gly Ala His Asn Ile Lys Glu Gln  62
RMCPII  Ala His Cys    Lys Gly Arg Glu Ile Thr Val Ile Leu    Gly Ala His Asp Val Arg Lys Arg  63
CA COW  Ala His Cys    Gly Val Thr Thr Ser Asp Val Val Val Ala Gly Glu Phe Asp Gln Gly Ser Ser  76
TR COW  Ala His Cys Tyr Lys Ser Gly Ile    Gln Val Arg Leu    Gly Gln Asp Asn Ile Asn Val Val  64

CCPI    Glu Lys Thr Gln Gln Val Ile Pro Met Val Lys Cys Ile Pro His Pro Asp Tyr Asn Pro Lys Thr  84
RMCPII  Glu Ser Thr Gln Gln Lys Ile Lys Val Glu Lys Gln Ile Ile His Glu Ser Tyr Asn Ser Val Pro  85
CA COW  Ser Glu Lys Ile Gln Lys Leu Lys Ile Ala Lys Val Phe Lys Asn Ser Lys Tyr Asn Ser Leu Thr  98
TR COW  Glu Gly Asn Gln Gln Phe Ile Ser Ala Ser Lys Ser Ile Val His Pro Ser Tyr Asn Ser Asn Thr  86

CCPI    Phe Ser Asn Asp Ile Met Leu Leu Lys Leu Lys Ser Lys Ala Lys    Arg Thr Arg Ala Val Arg  105
RMCPII  Asn Leu His Asp Ile Met Leu Leu Lys Leu Glu Lys Lys Val Glu    Leu Thr Pro Ala Val Asn  106
CA COW  Ile Asn Asn Asp Ile Thr Leu Leu Lys Leu Ser Thr Ala Ala Ser    Phe Ser Gln Thr Val Ser  119
TR COW  Leu Asn Asn Asp Ile Met Leu Ile Lys Leu Lys Ser Ala Ala Ser Leu Asn Ser Arg Val Ala     107
```

FIG. 4A

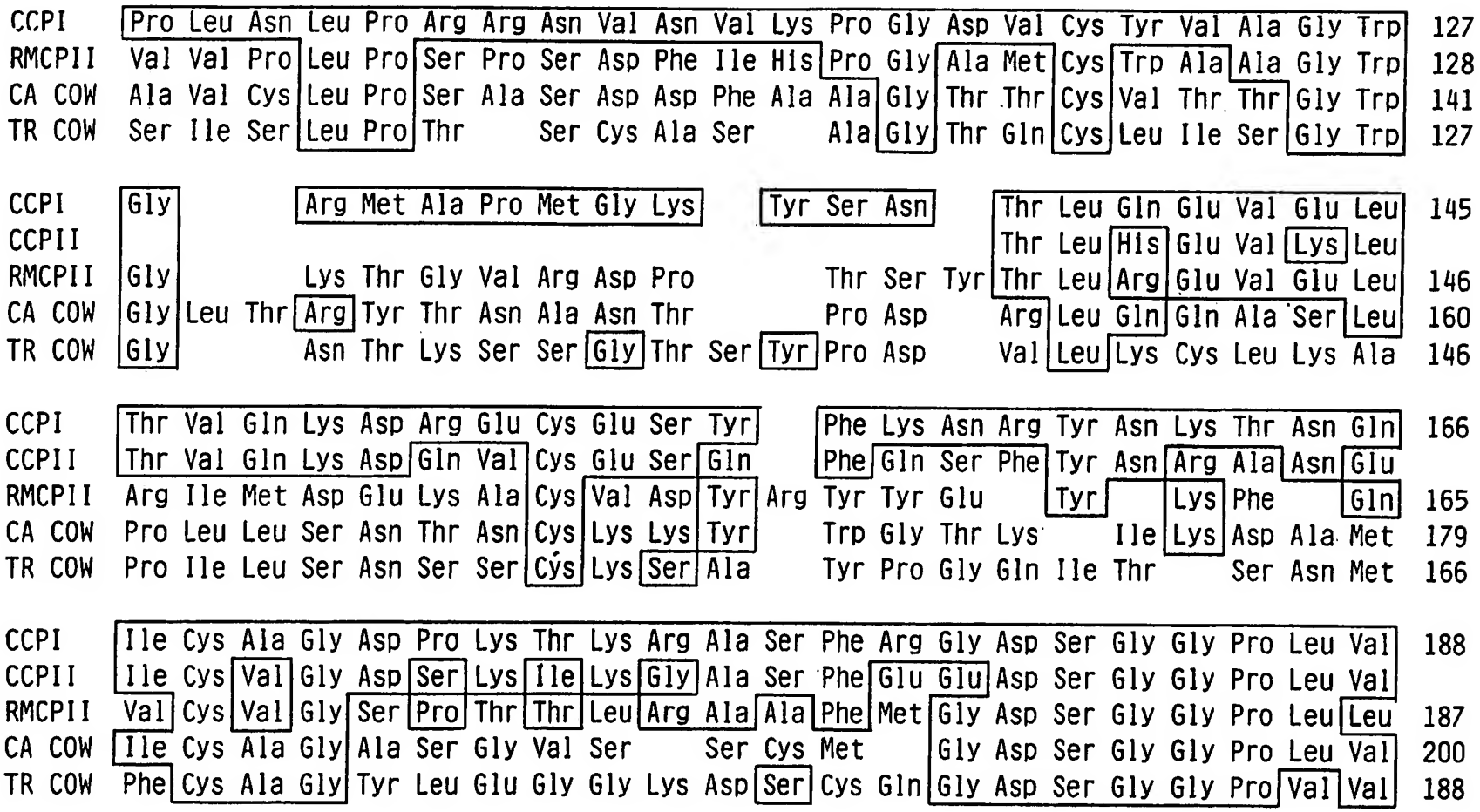

CCPI Pro Leu Asn Leu Pro Arg Arg Asn Val Asn Val Lys Pro Gly Asp Val Cys Tyr Val Ala Gly Trp 127
RMCPII Val Val Pro Leu Pro Ser Pro Ser Asp Phe Ile His Pro Gly Ala Met Cys Trp Ala Ala Gly Trp 128
CA COW Ala Val Cys Leu Pro Ser Ala Ser Asp Asp Phe Ala Ala Gly Thr Thr Cys Val Thr Thr Gly Trp 141
TR COW Ser Ile Ser Leu Pro Thr Ser Cys Ala Ser Ala Gly Thr Gln Cys Leu Ile Ser Gly Trp 127
CCPI Gly Arg Met Ala Pro Met Gly Lys Tyr Ser Asn Thr Leu Gln Glu Val Glu Leu 145
CCPII Thr Leu His Glu Val Lys Leu
RMCPII Gly Lys Thr Gly Val Arg Asp Pro Thr Ser Tyr Thr Leu Arg Glu Val Glu Leu 146
CA COW Gly Leu Thr Arg Tyr Thr Asn Ala Asn Thr Pro Asp Arg Leu Gln Gln Ala Ser Leu 160
TR COW Gly Asn Thr Lys Ser Ser Gly Thr Ser Tyr Pro Asp Val Leu Lys Cys Leu Lys Ala 146
CCPI Thr Val Gln Lys Asp Arg Glu Cys Glu Ser Tyr Phe Lys Asn Arg Tyr Asn Lys Thr Asn Gln 166
CCPII Thr Val Gln Lys Asp Gln Val Cys Glu Ser Gln Phe Gln Ser Phe Tyr Asn Arg Ala Asn Glu
RMCPII Arg Ile Met Asp Glu Lys Ala Cys Val Asp Tyr Arg Tyr Tyr Glu Tyr Lys Phe Gln 165
CA COW Pro Leu Leu Ser Asn Thr Asn Cys Lys Lys Tyr Trp Gly Thr Lys Ile Lys Asp Ala Met 179
TR COW Pro Ile Leu Ser Asn Ser Ser Cys Lys Ser Ala Tyr Pro Gly Gln Ile Thr Ser Asn Met 166
CCPI Ile Cys Ala Gly Asp Pro Lys Thr Lys Arg Ala Ser Phe Arg Gly Asp Ser Gly Gly Pro Leu Val 188
CCPII Ile Cys Val Gly Asp Ser Lys Ile Lys Gly Ala Ser Phe Glu Glu Asp Ser Gly Gly Pro Leu Val
RMCPII Val Cys Val Gly Ser Pro Thr Thr Leu Arg Ala Ala Phe Met Gly Asp Ser Gly Gly Pro Leu Leu 187
CA COW Ile Cys Ala Gly Ala Ser Gly Val Ser Ser Cys Met Gly Asp Ser Gly Gly Pro Leu Val 200
TR COW Phe Cys Ala Gly Tyr Leu Glu Gly Gly Lys Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Val Val 188

FIG. 4B

```
CCPI     Cys Lys Lys Val Ala Ala                 Gly Ile Val Ser Tyr Gly Tyr Lys Asp             Gly  204
CCPII    Cys Lys Arg Ala Ala Ala                 Gly Ile Val Ser Tyr Gly Gln Thr Asp             Gly
RMCPII   Cys Ala Gly Val Ala His                 Gly Ile Val Ser Tyr Gly His Pro Asp             Ala  203
CA COW   Cys Lys Lys Asn Gly Ala Trp Thr Leu Val Gly Ile Val Ser Trp Gly Ser Ser Thr Cys Ser Thr  222
TR COW   Cys Ser Gly Lys Leu Gln                 Gly Ile Val Ser Trp Gly Ser     Gly Cys Ala Gln  205

CCPI     Ser Pro     Pro Arg Ala Phe Tyr Lys Val Ser Ser Phe Leu Ser Trp Ile Lys Lys Thr Met Lys  225
CCPII    Ser Ala     Pro Gln Val Phe Thr Arg Val Leu Ser Phe Val Ser Trp Ile Lys Lys Thr Met Lys
RMCPII   Lys Pro     Pro Ala Ile Phe Thr Arg Val Ser Thr Tyr Val Pro     Thr Ile Asn Ala Val Ile  223
CA COW   Ser Thr     Pro Gly Val Tyr Ala Arg Val Thr Ala Leu Val Asn Trp Val Gln Gln Thr Leu Ala  243
TR COW   Lys Asn Lys Pro Gly Val Tyr Thr Lys Val Cys Asn Tyr Val Ser Trp Ile Lys Gln Thr Ile Ala  227

CCPI     Ser Ser  227
CCPII    His Ser
RMCPII   Asn      224
CA COW   Ala Asn  245
TR COW   Ser Asn  229
```

FIG. 4C

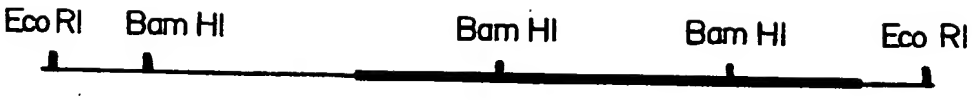
```
        10         20         30         40         50         60         70         80         90        100
ACCTCCTTCCTCCCCTTCCAAGGGCTTTGTCTACTCTGCCCTCCTCCATCCTGAGCAGCCTTCCTGGGAAGATGCCACCAGTCCTGATTCTCCTGACCCT
       110        120        130        140        150        160        170        180        190        200
ACTTCTGCCTCTCAGAGCTGGAGCAGAGGAGATAATCGGAGGCAATGAGATCAGTCCACATTCCCGTCCCTACATGGCATATTATGAGTTTCTGAAAGTT
       210        220        230        240        250        260        270        280        290        300
GGTGGGAAGAAGATGTTCTGCGGAGGCTTCCTGGTTCGAGACAAATTCGTGCTAACAGCTGCTCACTGCAAAGGAAGCTCAATGACAGTCACACTGGGGG
       310        320        330        340        350        360        370        380        390        400
CTCACAACATCAAGGCTAAGGAGGAGACACAGCAGATCATCCCTGTGGCAAAAGCCATTCCCCATCCAGACTATAATCCTGATGACCGTTCTAATGACAT
       410        420        430        440        450        460        470        480        490        500
CATGCTATTAAAGCTGGTGAGAAATGCCAAGAGGACTAGAGCTGTGAGGCCCCTCAACCTGCCCAGGCGCAATGCTCATGTGAAGCCAGGGGATGAGTGC
       510        520        530        540        550        560        570        580        590        600
TATGTGGCTGGTTGGGGAAAGGTAACCCCGGACGGGGAATTCCCAAAAACACTGCACGAAGTTAAGCTGACAGTACAGAAGGATCAGGTGTGTGAGTCCC
       610        620        630        640        650        660        670        680        690        700
AGTTCCAAAGTTCTTACAACAGAGCTAATGAGATATGTGTGGGAGACTCAAAGATCAAGGGAGCTTCCTTTGAGGAGGATTCTGGAGGCCCGCTTGTGTG
       710        720        730        740        750        760        770        780        790        800
TAAAAGAGCAGCTGCAGGCATCGTCTCCTACGGGCAAACTGATGGATCAGCTCCGCAAGTCTTCACAAGAGTTTTGAGTTTTGTATCGTGGATAAAGAAA
       810        820        830        840        850        860        870        880        890        900
ACGATGAAACACAGCTAACTACAAGAAGCAACTAGATCCTGACTGACAGCCATCTTCCCATAGCTGAGTCCAGGATTGCTCTAGGACAGATGGCAGGCAA
       910        920        930        940
CTGAATAAAGAACTTTCTCTGACTGCAAAAAAAAAAAAA
```

FIG. 5

```
        10        20        30        40        50        60        70        80        90       100
CGCAATCCTGCTTCTGCTGGCCTTCCTCCTGCTGCCCAGGGCAGATGCAGGGGAGATCATCGGGGGACATGAGGCCGAGCCCCACTCCCGCCCCTACATG
       110       120       130       140       150       160       170       180       190       200
GCTTATCTTATGATCTGGGATCAGAAGTCTCTGAAGAGGTGCGGTGGCTTCCTGATACAAGACGACTTCGTGCTGACAGCTGCTCACTGTTGGGGAAGCT
       210       220       230       240       250       260       270       280       290       300
CCATAAATGTCACCTTGGGGGCCCACAATATCAAAGAACAGGAGCCGACCCAGCAGTTTATCCCTGAAAAGACCCATCCCCCATCCAGCCTATAATCCTA
       310       320       330       340       350       360       370       380       390       400
AGAACTTCTCCAACGACATCATGCTACTGCAGCTGGAGAGAAAGGCCAAGCGGACCAGAGCTGTGCAGCCCCTCAGGCTACCTAGCAACAAGGCCCAGGT
       410       420       430       440       450       460       470       480       490       500
GAAGCCAGGGCAGACATGCAGTGTGGCCGGCTGGGGGCAGACGGCCCCCCTGGGAAAACACTCACACACACTACAAGAGGTGAAGATGACAGTGCAGGAA
       510       520       530       540       550       560       570       580       590       600
GATCGAAAGTGCGAATCTGACTTACGCCATTATTACGACAGTACCATTGAGTTGTGCGTGGGGGACCCAGAGATTAAAAAGACTTCCTTTAAGGGGGACT
       610       620       630       640       650       660       670       680       690       700
CTGGAGGCCTCTTGTGTGTAACAAGGTGGCCCAGGGCATTGTCTCCTATGGACGAAACAATGGCATGCCTCCACGAGCCTGCACCAAAGTCTCAAGCTTT
       710       720       730       740       750       760       770       780       790       800
GTACACTGGATAAAGAAAACCATGAAAGGCTACTAACTACAGGAAGCAAACTAAGCCCCCGCTGTAATGAAACACCTTCTCTGGAGCCAAGTCCAGATTT
       810       820       830       840
ACACTGGGAGAGGTGCCAGCAACTGAATAAATACCTCTTAGCTGAGTGG
```

FIG. 6

FIG. 7



```
        10         20         30         40         50         60         70         80         90        100
ATGCAGCCATTCCTCCTCCTGTTGGCCTTTCTTCTGACCCCTGGGGCTGGGACAGGTAAGTGACTATCCCTATTCCAGAGGCCTGAACCCATCTTATAAG
       110        120        130        140        150        160        170        180        190        200
ATACCCTGTACCCATGAGCACTGGTCAGGAATTTTCCTCAATCTGAGCCCACCTCCCATTCCACACCCAGACTTATAAATCTGAGGCTAGATAGACACTC
       210        220        230        240        250        260        270        280        290        300
AGCAAAGATCGAAAAATGAAGGGTGTTCCCTAAAAGGTTTAATGGGTGTTAGCCTCTCCCTAGACCTCTCCTTTATGACCTGGAGTGTGGATTGTTCTTA
       310        320        330        340        350        360        370        380        390        400
GAAAGGCATTTGGTAGGGAATGTGAAGCTAAAAAAGATAAGTAATTATTACTCTACACTCCAACCCAGGAAAGAGGAGCTCAGACACCAAGTTGCAGTCA
       410        420        430        440        450        460        470        480        490        500
TAGGAGTTGTTACTGGACTCAGCTTAAGAACCACTTATTCAGTGCCCCCACCCAGTCACCCCTGCCAGGAGGGAAGCCCATGGTGCAACTGATCTCAGAG
       510        520        530        540        550        560        570        580        590        600
ATGGCAAGATGACTGTGTCCATAGCTCTCCCATTACCTTGGCTCCACCTGGGCTTTGCGATTCATTTTTAGTTGATTTCTCCACTTCCTTCTGCCTTTGC
       610        620        630        640        650        660        670        680        690        700
AGGCCCTCAGCTACTCCACTGACCTGGTGATAACCCCCTCTAACATCCCTGAGGTCCTGAATCCCACCAGCACTACCCCCACTAAACCTCAGCCAAAGGC
       710        720        730        740        750        760        770        780        790        800
TAATTGGAGGCTATTCATTTATGCACCAAACAACACTTACTGAGAACCTAGAATGTGTTCAGCCCTGGCACATGAGAATTTTAGAAAATCCAACTCCAGA
       810        820        830        840        850        860        870        880        890        900
AGCTCATGGGTAGATCATTAGAAAATGGCACCAATCAGGAGGACAGCAGAGGCTCAGAGGAATCAGAGAGACTCACTAAAAAGGAGGAGCCTGGGAAGTC
       910        920        930        940        950        960        970        980        990       1000
CTGTCCAGGATCCTCTGGGGCTCTACCACTGTGAAAATAGATCTTTCTTCCTCCATGTGTGTTTCTGCAGTAGGATGCAGAAGTCAAGCACTGTCCATTC
```

FIG. 8A

```
      1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
TTTCTGAGAGAAGGTTCTGAGACTTAAAGTTCAGAAAAGCTTGCTCTGCCTGCAAGGAGTACTGCCTGCAGAGCTCAGGAGGACCCCTCCCTCACTGTCT

      1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
GCAGCACACAAAACACTTCACCCTCAATCACTGTCGTCCTAAGCCTTCTCTCCTGACTGCAGCCATCCCCAGCTACTTTATCTTCCAGTCTTTCCTTTCA

      1210       1220       1230       1240       1250       1260       1270       1280       1290       1300
GAGGAGATCATCGGGGGCCATGAGGCCAAGCCCCACTCCCGCCCCTACATGGCCTTTGTTCAGTTTCTGCAAGAGAAGAGTCGGAAGAGGTGTGGCGGCA

      1310       1320       1330       1340       1350       1360       1370       1380       1390       1400
TCCTAGTGAGAAAGGACTTTGTGCTGACAGCTGCTCACTGCCAGGGAAGGTAAGGAGCAGCACCAGCTCACCTCCTGAGTTCCCCCTACAGGGACCCTTG

      1410       1420       1430       1440       1450       1460       1470       1480       1490       1500
TTTTCTCCTGGGGACTGCAGCCCAGGGGAGCTTCCAGAGTTCCTGGTAACACAAGCCCATGAAAGCTCATCGGAGCAGCCACCTGGGAATAGGACTAAGA

      1510       1520       1530       1540       1550       1560       1570       1580       1590       1600
TGATGAAAGGCTGAGAAATAGGACAGGGTCAGAGGGTCAGAGGGTCAGAGGTCAGAGCAAGTGGCATGGTTTCACACACTGGCCCAATAAGGCAGTAGTA

      1610       1620       1630       1640       1650       1660       1670       1680       1690       1700
AAGGTTGAATCCATGCATTAAGGATCAAAATGCAAACTCGAGCAATTTCATGATATTTCCTGAAGGCAAGAGAGAGCCCCCAGAAGCCCTACTCTTCATG

      1710       1720       1730       1740       1750       1760       1770       1780       1790       1800
TCTTTGTGAAGTAGAGGGCACAGTCACTCAGCCCTGGAGCCCTCCTGTCCCTTCAACTTCCCCTCAGCTGCAGCCCTACCCTGCCACAGCTGTCATCTGC

      1810       1820       1830       1840       1850       1860       1870       1880       1890       1900
TCTTCACTGCTCCTGGGCTCTATCCCCTGTGACTCCACCCCCATCCTCACTCTGCTCTCTGTGCAGCTCCATAAATGTCACGTTGGGGGCCCACAATATC

      1910       1920       1930       1940       1950       1960       1970       1980       1990       2000
AAGGAACAGGAGCGGACCCAGCAGTTTATCCCTGTGAAAAGACCCATCCCCCATCCAGCCTATAATCCTAAGAACTTCTCCAACGACATCATGCTACTGC
```

FIG. 8B

```
    2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
AGGTGAGGCACACTCCTGCCACTCTTGCTCTTCTTGGTCCAGTTGGTTCCACTCCCACTGGGATGCCGGCCCTTTCCTCCTTTCCATCCTGACCTCTTGG
    2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
TCAGTTCCTGTGCCTTAGAGGAGAGGGAAGATTGTGCAGCCCCATCACTGTGTCGGGGCCCAGAAGGCCATTGCCTGACCTGGACTTTCTTGCTTCTTCC
    2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
CCACCAGCTGGAGAGAAAGGCCAAGTGGACCACAGCTGTGCGGCCTCTCAGGCTACCTAGCAGCAAGGCCCAGGTGAAGCCAGGGCAGCTGTGCAGTGTG
    2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
GCTGGCTGGGGTTATGTCTCAATGAGCACTTTAGCAACCACACTGCAGGAAGTGTTGCTGACAGTGCAGAAGGACTGCCAGTGTGAACGTCTCTTCCATG
    2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
GCAATTACAGCAGAGCCACTGAGATTTGTGTGGGGGATCCAAAGAAGACACAGACCGGTTTCAAGGTTGGGTTTCCAGCCTCTACCCAGAAAGACCACAG
    2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
GGAGAGAGGAACTTGAGGGAGTTCTGGGGTATGACAGTGGCCAGATCTTTATGCTCTCAGCCCAGAGCTTGGGCAGCCTGTGTCCCCCTGGAATCTAGTC
    2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
TTTAGTCCACTCCTTGGGTGGACTGGGGTGGCTGGAGGGGAAGGATCTGTCATCCGACAGTCACTTATCACCAAAGTGTCCGTGAGAATTGCAGAGAAAG
    2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
GCTTGGTCTGGGGAGAGCCAGAAAGCACAGCCAGGGCCATGCTGGAGACCCAGGACTGAGGGAGGTGAGTGACAAGGCCTGCCTGGCACTGATGTCACAC
    2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
AAAATCCATAACAGAGACCACCCATCCCAGGGAACACTAGCTCAGTCTTTCCTCTCTCTGTTCACAGGGGGACTCCGGGGGGCCCCTCGTGTGTAAGGAC
    2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
GTAGCCCAAGGTATTCTCTCCTATGGAAACAAAAAAGGGACACCTCCAGGAGTCTACATCAAGGTCTCACACTTCCTGCCCTGGATAAAGAGAACAATGA
    3010
AGCGCCTCTAA
```

FIG. 8C

```
                                                30
ATG CAG CCA TTC CTC CTC CTG TTG GCC TTT CTT CTG ACC CCT GGG
Met Gln Pro Phe Leu Leu Leu Ala Phe Leu Leu Thr Pro Gly

                    60                                  90
GCT GGG ACA GAG GAG ATC ATC GGG GGC CAT GAG GCC AAG CCC CAC
Ala Gly Thr Glu Glu Ile Ile Gly Gly His Glu Ala Lys Pro His

                                               120
TCC CGC CCC TAC ATG GCC TTT GTT CAG TTT CTG CAA GAG AAG AGT
Ser Arg Pro Tyr Met Ala Phe Val Gln Phe Leu Gln Glu Lys Ser

                   150                                  180
CGG AAG AGG TGT GGC GGC ATC CTA GTG AGA AAG GAC TTT GTG CTG
Arg Lys Arg Cys Gly Gly Ile Leu Val Arg Lys Asp Phe Val Leu

                                               210
ACA GCT GCT CAC TGC CAG GGA AGC TCC ATA AAT GTC ACC TTG GGG
Thr Ala Ala His Cys Gln Gly Ser Ser Ile Asn Val Thr Leu Gly

                   240                                  270
GCC CAC AAT ATC AAG GAA CAG GAG CGG ACC CAG CAG TTT ATC CCT
Ala His Asn Ile Lys Glu Gln Glu Arg Thr Gln Gln Phe Ile Pro

                                               300
GTG AAA AGA CCC ATC CCC CAT CCA GCC TAT AAT CCT AAG AAC TTC
Val Lys Arg Pro Ile Pro His Pro Ala Tyr Asn Pro Lys Asn Phe

                   330                                  360
TCC AAC GAC ATC ATG CTA CTG CAG CTG GAG AGA AAG GCC AAG TGG
Ser Asn Asp Ile Met Leu Leu Gln Leu Glu Arg Lys Ala Lys Trp
```

FIG. 9A

```
                                                    390
ACC ACA GCT GTG CGG CCT CTC AGG CTA CCT AGC AGC AAG GCC CAG
Thr Thr Ala Val Arg Pro Leu Arg Leu Pro Ser Ser Lys Ala Gln

                420                                     450
GTG AAG CCA GGG CAG CTG TGC AGT GTG GCT GGC TGG GGT TAT GTC
Val Lys Pro Gly Gln Leu Cys Ser Val Ala Gly Trp Gly Tyr Val

                                                    480
TCA ATG AGC ACT TTA GCA ACC ACA CTG CAG GAA GTG TTG CTG ACA
Ser Met Ser Thr Leu Ala Thr Thr Leu Gln Glu Val Leu Leu Thr

                510                                     540
GTG CAG AAG GAC TGC CAG TGT GAA CGT CTC TTC CAT GGC AAT TAC
Val Gln Lys Asp Cys Gln Cys Glu Arg Leu Phe His Gly Asn Tyr

                                                    570
AGC AGA GCC ACT GAG ATT TGT GTG GGG GAT CCA AAG AAG ACA CAG
Ser Arg Ala Thr Glu Ile Cys Val Gly Asp Pro Lys Lys Thr Gln

                600                                     630
ACC GGT TTC AAG GGG GAC TCC GGG GGG CCC CTC GTG TGT AAG GAC
Thr Gly Phe Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Lys Asp

                                                    660
GTA GCC CAA GGT ATT CTC TCC TAT GGA AAC AAA AAA GGG ACA CCT
Val Ala Gln Gly Ile Leu Ser Tyr Gly Asn Lys Lys Gly Thr Pro

                690                                     720
CCA GGA GTC TAC ATC AAG GTC TCA CAC TTC CTG CCC TGG ATA AAG
Pro Gly Val Tyr Ile Lys Val Ser His Phe Leu Pro Trp Ile Lys

AGA ACA ATG AAG CGC CTC TAA
Arg Thr Met Lys Arg Leu End
```

FIG. 9B

```
1  Met Gln Pro Phe Leu Leu Leu Leu Ala Phe Leu Leu Thr Pro Gly Ala Gly Thr Glu Glu
   Ile Ile Gly Gly His Glu Ala Lys Pro His Ser Arg Pro Tyr Met Ala Phe Val Gln
1  Met Lys Ile Leu Leu Leu Leu Leu Thr Leu Ser Leu Ala Ser Arg Thr Lys Ala Gly Glu
   Ile Ile Gly Gly His Glu Val Lys Pro His Ser Arg Pro Tyr Met Ala Leu Leu Ser
1  Met Pro Pro Val Leu Ile Leu Leu Thr Leu Leu Leu Pro Leu Arg Ala Gly Ala Glu Glu
   Ile Ile Gly Gly Asn Glu Ile Ser Pro His Ser Arg Pro Tyr Met Ala Tyr Tyr Glu Phe
1  Met Pro Pro Val Leu Ile Leu Leu Thr Leu Leu Leu Pro Leu Gly Ala Gly Ala Glu Glu
   Ile Ile Gly Gly His Val Val Lys Pro His Ser Arg Pro Tyr Met Ala Phe Val Lys Ser
1  Met Pro Pro Ile Leu Ile Leu Leu Thr Leu Leu Leu Pro Leu Arg Ala Gly Ala Glu Glu
   Ile Ile Gly Gly His Glu Val Lys Pro His Ser Arg Pro Tyr Met Ala Arg Val Arg Phe

40 Phe Leu Gln Glu Lys Ser Arg Lys Arg Cys Gly Gly Ile Leu Val Arg Lys Asp Phe Val
   Leu Thr Ala Ala His Cys Gln Gly Ser Ser Ile Asn Val Thr Leu Gly Ala His Asn Ile
40 Ile Lys Asp Gln Gln Pro Glu Ala Ile Cys Gly Gly Phe Leu Ile Arg Glu Asp Phe Val
   Leu Thr Ala Ala His Cys Glu Gly Ser Ile Ile Asn Val Thr Leu Gly Ala His Asn Ile
41 Leu Lys Val Gly Gly Lys Lys Met Phe Cys Gly Gly Phe Leu Val Arg Asp Lys Phe Val
   Leu Thr Ala Ala His Cys Lys Gly Ser Ser Met Thr Val Thr Leu Gly Ala His Asn Ile
41 Val Asp Ile Glu Gly Asn Arg Arg Tyr Cys Gly Gly Phe Leu Val Gln Asp Asp Phe Val
   Leu Thr Ala Ala His Cys Arg Asn Arg Thr Met Thr Val Thr Leu Gly Ala His Asn Ile
41 Val Lys Asp Asn Gly Lys Arg His Ser Cys Gly Gly Phe Leu Val Gln Asp Tyr Phe Val
   Leu Thr Ala Ala His Cys Thr Gly Ser Ser Met Arg Val Ile Leu Gly Ala His Asn Ile

80 Lys Glu Gln Glu Arg Thr Gln Gln Pbe Ile Pro Val Lys Arg Pro Ile Pro His Pro Ala
   Tyr Asn Pro Lys Asn Phe Ser Asn Asp Ile Met Leu Leu Gln Leu Glu Arg Lys Ala Lys
80 Lys Glu Gln Glu Lys Thr Gln Gln Val Ile Pro Met Val Lys Cys Ile Pro His Pro Asp
   Tyr Asn Pro Lys Thr Phe Ser Asn Asp Ile Met Leu Leu Lys Leu Lys Ser Lys Ala Lys
81 Lys Ala Lys Glu Glu Thr Gln Gln Ile Ile Pro Val Ala Lys Ala Ile Pro His Pro Asp
   Tyr Asn Pro Asp Asp Arg Ser Asn Asp Ile Met Leu Leu Lys Leu Val Arg Asn Ala Lys
81 Lys Ala Lys Glu Glu Thr Gln Gln Ile Ile Pro Val Ala Lys Ala Ile Pro His Pro Asp
   Tyr Asn Ala Thr Ala Phe Phe Ser Asp Ile Met Leu Leu Lys Leu Glu Ser Lys Ala Lys
81 Arg Ala Lys Glu Glu Thr Gln Gln Ile Ile Pro Val Ala Lys Ala Ile Pro His Pro Ala
   Tyr Asp Asp Lys Asp Asn Thr Ser Asp Ile Met Leu Leu Lys Leu Glu Ser Lys Ala Lys
```

FIG. 10A

```
120 Trp Thr Thr Ala Val Arg Pro Leu Arg Leu Pro Ser Ser Lys Ala Gln Val Lys Pro Gly
    Gln Leu Cys Ser Val Ala Gly Trp Gly     Tyr Val Ser Met Ser Thr Leu Ala Thr
120 Arg Thr Arg Ala Val Arg Pro Leu Asn Leu Pro Arg Arg Asn Val Asn Val Lys Pro Gly
    Asp Val Cys Tyr Val Ala Gly Trp Gly     Arg Met Ala Pro Met Gly Lys Tyr Ser Asn
121 Arg Thr Arg Ala Val Arg Pro Leu Asn Leu Pro Arg Arg Asn Ala His Val Lys Pro Gly
    Asp Glu Cys Tyr Val Ala Gly Trp Gly     Lys Val Thr Pro Asp Gly Glu Phe Pro Lys
121 Arg Thr Lys Ala Val Arg Pro Leu Lys Leu Pro Arg Pro Asn Ala Arg Val Lys Pro Gly
    Asp Val Cys Ser Val Ala Gly Trp Gly Ser Arg Ser Ile Asn Asp Thr Lys Ala Ser Ala
121 Arg Thr Lys Ala Val Arg Pro Leu Lys Leu Pro Arg Pro Asn Ala Arg Val Lys Pro Gly
    His Val Cys Ser Val Ala Gly Trp Gly Arg Thr Ser Ile Asn Ala Thr Gln Arg Ser Ser

158 Thr Leu Gln Glu Val Leu Leu Thr Val Gln Lys Asp Cys Gln Cys Glu Arg Leu Phe His
    Gly Asn Tyr Ser Arg Ala Thr Glu Ile Cys Val Gly Asp Pro Lys Lys Thr Gln Thr Gly
159 Thr Leu Gln Glu Val Ala Leu Thr Val Gln Lys Asp Arg Glu Cys Glu Ser Tyr Phe Lys
    Asn Arg Tyr Asn Lys Thr Asn Gln Ile Cys Ala Gly Asp Pro     Lys Thr Lys Arg Ala
160 Thr Leu His Glu Val Lys Leu Thr Val Gln Lys Asp Gln Val Cys Glu Ser Gln Phe Gln
    Ser Ser Tyr Asn Arg Ala Asn Glu Ile Cys Val Gly Asp Ser Lys Ile Lys Gly Ala
161 Arg Leu Arg Glu Ala Gln Leu Val Ile Gln Glu Asp Glu Cys Lys Lys Arg Phe Arg
    His     Tyr Thr Glu Thr Thr Glu Ile Cys Ala Gly Asp Leu Lys Ile Lys Thr Pro
161 Cys Leu Arg Glu Ala Gln Leu Ile Gln Lys Asp Lys Glu Cys Lys Lys Tyr Phe Tyr
    Lys Tyr Phe Lys Thr Met Gln Ile Cys Ala Gly Asp Pro Lys Ile Gln Ser Thr Leu

198     Phe Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Lys Asp Val Ala Gln Gly Ile Leu
    Ser Tyr Gly Asn Lys Lys Gly Thr Pro Pro Gly Val Tyr Ile Lys Val Ser His Phe Leu
198 Ser Phe Arg Gly Asp Ser Gly Gly Pro Leu Val Cys Lys Lys Val Ala Ala Gly Ile Val
    Ser Tyr Gly Tyr Lys Asp Ser Gly Pro Pro Arg Ala Phe Thr Lys Val Ser Ser Phe Leu
199 Ser Phe Glu Glu Asp Ser Gly Gly Pro Leu Val Cys Lys Arg Ala Ala Gly Ile Val
    Ser Tyr Gly Gln Thr Asp Gly Gly Ser Ala Pro Gln Val Cys Phe Thr Arg Val Leu Ser Phe Val
200     Phe Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Asp Asn Lys Ala Tyr Gly Leu Leu
    Ala Tyr Ala Lys Asn Arg Thr Ile Ser Ser Gly Val Phe Thr Lys Ile Val His Phe Leu
200     Tyr Ser Gly Asp Ser Gly Gly Pro Leu Val Cys Asn Asn Lys Ala Tyr Gly Val Leu
    Thr Tyr Gly Leu Asn Arg Thr Ile Gly Pro Gly Val Phe Thr Lys Val Val His Tyr Leu

237 Pro Trp Ile Lys Arg Thr Met Lys Arg Leu
238 Ser Trp Ile Lys Lys Thr Met Lys Ser Ser
239 Ser Trp Ile Lys Lys Thr Met Lys His Ser
239 Pro Trp Ile Ser Arg Asn Met Lys Leu Leu
239 Pro Trp Ile Ser Arg Asn Met Lys Leu Leu
```

FIG. 10B

EF2394	Asp-Val-Asp-Ala
EF2395	Ala-Pro-Asp-Ala
EF2396	Ala-Asn-Pro-Ala
EF2397	Phe-Pro-Arg-Phe
EF2398	Ala-Pro-Arg-Phe
EF2368	Phe-Pro-Asp-Phe
EF2369	Phe-Pro-Asn-Phe
EF2372	Phe-Asn-Pro-Phe
EF2373	Phe-Asp-Pro-Phe

Ala = alanine
Arg = arginine
Asn = asparagine
Asp = aspartic acid
Phe = phenylalanine
Pro = proline
Val = valine

FIG. 11

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/00340

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): C 07K 13/00; C12N 9/48, 9/64
U.S. CL: 530/300; 435/226, 212;935/23

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System	Classification Symbols
U.S.	530/300; 435/226,212; 935/23

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [8]

DIALOG

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
X	Cell, Vol.49, issued 05 June 1987, D. Masson et al, "A Family of Serine Esterases in Lytic Granules of Cytolytic T Lymphocytes", pages 679-685, see entire document.	1
X	FEBS Letters, vol. 234, number 1, issued July 1988, R.Bleackley et al, "Isolation of two cDNA sequences which encode cytotoxic cell proteases", pages 153-159, see entire document.	1
X	Proteins: Structure, Function, and Genetics, volume 4, issued 1988, M.Murphy et al, "Comparative Molecular Model Building of Two Serine Proteinases From Cytotoxic T Lymphocytes, pages 190-204, see entire document.	1

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
12 APRIL 1991	24 MAY 1991

International Searching Authority	Signature of Authorized Officer
ISA/US	KEITH HENDRICKS

Form PCT/ISA/210 (second sheet) (rev 11-87)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
X	Biochemistry, vol. 27, No.18, issued 1988, C.Kobe et al, "Organization of two Genes Encoding Cytotoxic T Lymphocyte-Specific Serine Proteases CCPI and CCPII", pages 6941-6946, see entire document.	2

FURTHER INFORMA. .N CONTINUED FROM THE SECOND SHEET

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers , because they relate to subject matter [2] not required to be searched by this Authority, namely:

2. ☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [3], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☒ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

Invention I: claim 1, a vector with CCP1.
Invention II:claim 2, a vector with CCP2.
Invention III: claim 3, a vector with hCCP1.
Invention IV: claim 4, a vector with hCCPX. see attachment

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers: 1

Telephone practice

4 ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (Supplemental Sheet (1)) (July 1992)

Continued from sheet 2 (supplemental)

Invention V: claim 5, a CCP1 protein
Invention VI: claim 6, a CCP2 protein
Invention VII: claim 7, a hCCP1 protein
Invention VIII: claim 8, a hCCPX protein
Invention IX: claim 9, an inhibitory peptide.

Each of the inventions I-IX are distinct from the other as separate products which fail to meet the criteria of PCT Rules 13.1 and 13.2 regarding unity of invention.